UNCLASSIFIED

AD 415178

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

63-4-5

REPORT NO. 1291262

PROJECT TRIDENT

TECHNICAL REPORT

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

ARTHUR D. LITTLE, INC.
35 ACORN PARK CAMBRIDGE, MASSACHUSETTS

DEPARTMENT OF THE NAVY
BUREAU OF SHIPS

N0bsr-81564 S-7001-0307

DECEMBER 1962

REPORT NO. 1291262

PROJECT TRIDENT

TECHNICAL REPORT

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

by

Gordon Raisbeck

ARTHUR D. LITTLE, INC.

35 ACORN PARK CAMBRIDGE, MASSACHUSETTS

DEPARTMENT OF THE NAVY

BUREAU OF SHIPS

N0bsr-81564 S-7001-0307

DECEMBER 1962

PREFACE

This is one of a series of technical reports being issued by Arthur D. Little, Inc., under Contract NObsr-81564 with the Bureau of Ships as part of the Project TRIDENT.

FOREWORD

The object of this report is to explain some of the ideas in modern information theory and to show how they can be applied to certain problems in signal transmission and signal detection. It is not intended as a text or reference work. It evolved from several sets of lectures at various times and places to audiences of scientists and engineers who had no specialized knowledge of communications or information theory. The earliest sections, which introduce the fundamental ideas of amount of information and channel capacity, may nevertheless be of interest to readers with less technical background.

We thank the Institute for Defense Analyses for permission to use herein portions of IDA Technical Note 60-19, "Modulation, Coding and Information Theory," which was written with the support of Contract NOSD-50 with the Advance Research Projects Agency; our colleagues J. Kaiser, G. Sutton, and others at IDA, who heard and criticized a series of lectures on which the Technical Note was based; our colleagues at Bell Telephone Laboratories, Inc., and Arthur D. Little, Inc., who likewise criticized subsequent oral presentations; Hugh Leney, M. S. Klein, and Paul B. Coggins of A. D. Little, Inc., and Professor John Wozencraft of M.I.T., who read and criticized the manuscript; and Claude E. Shannon, E. N. Gilbert, J. R. Pierce, C. C. Cutler, R. M. Fano and others from whose publications we have borrowed liberally.

ABSTRACT

This is an expository essay on information theory for engineers interested in communications, sonar, and radar who have no specialized knowledge of statistical communication theory. The fundamental concepts of information theory, and in particular, quantity of information and channel capacity, are defined and explained in simple terms. These concepts are used to make a quantitative estimate of the performance of several common modulation schemes and to analyze the performance of search and detection systems. The effectiveness of repeated or prolonged observations on detection thresholds and reliability of detection, and the relative performance of coherent and incoherent integration, are explained and illustrated quantitatively.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

I. INTRODUCTION

Some paradoxes and misunderstandings about information have arisen in recent years as the science of information theory has been disseminated. The first misunderstanding is the belief that any intelligent person ought to know what the word information means.

In any specialized study, new concepts arise which have to have names. Sometimes we name the concept after a person: Doppler shift, Plank's constant. Sometimes we give it a number or letter: The first law of thermodynamics, X-rays. Sometimes we make up a new word: Meson, radio. But often we use a common word: Current, mass.

When a new technical concept is named with a common word, the word acquires a new meaning. It is impossible to use the word in a technical context until that new meaning has been defined. Pressing a suit does not mean the same thing to a lawyer that it does to a tailor. And information does not mean the same thing to a communications engineer that it does to a police detective. There is no reason to expect anyone to know what the word information means to an information theorist unless he has been told.

In this report, we shall give the information theorist's definition of information, and some examples of how the word is used in its technical sense. In this way, we shall indicate why the concept is useful enough to be worth a name of its own, and attempt to show that the concept has enough in common with a nontechnical idea of information that no real violence is done to the language in appropriating this word to name it. Then we shall use the new concept as a tool to investigate the properties of certain communication systems and detection systems.

It is possible simply to state a mathematical definition of information, and proceed to demonstrate some of its properties. However, such an approach is likely to be unconvincing, because the definition itself does not indicate just why it was chosen. As an alternative, we shall discuss some reasonable and useful properties which we can hope a new definition of information will have, and use them to narrow down the search.

II. GENERALIZED COMMUNICATION SYSTEM

A generalized communication system is illustrated in Figure 1. The first element of this system is an information source. Although we have not yet defined what we mean by information, assume that the information source is a person talking. The output of the information source is called a message. If the information source is a person talking, the message is what he says.

The next element in the communication system is a transmitter. The transmitter transforms the message in some way and produces a signal suitable for transmission over the next element of this system, the communication channel. The input to the transmitter is the message, and the output of the transmitter is the signal. If the transmitter is a telephone handset, the signal is an electrical current proportional to the pressure of the sound waves impinging on the mouthpiece of the instrument.

The next element of this communication system is the channel. This is the medium used to transmit the signal from the transmitter to the receiver. While going through the channel, the signal may be altered by noise or distortion. In principle, noise and distortion may be differentiated on the basis that distortion is a fixed operation applied to the signal, while noise involves statistical and unpredictable perturbations. All or part of the effect of distortion can be corrected by applying the inverse operation or a partial inverse operation, but a perturbation due to noise cannot always be removed, because the signal does not always undergo the same change during transmission. In practice, the gamut of perturbation runs from noise to distortion. The input to the channel is the signal, sometimes called the transmitted signal. The output of the channel is the received signal, supposed to be in some sense a faithful representation of the transmitted signal.

The next element in this idealized communication system is the receiver. This operates on the received signal and attempts to reproduce from it the original message. It will ordinarily perform an operation which is approximately the inverse of the operation performed by the transmitter. The two operations may differ somewhat, however, because the receiver may also be required to combat the noise and distortion in the channel. The input to the receiver is the received signal, and the output of the receiver is the received message.

The last element of this communication system is the destination. This is the person or thing for whom the message is intended.

III. DEFINITION OF INFORMATION

An intuitively and esthetically desirable definition of amount of information will be a measure of time or cost of transmitting messages. When applied to a message source, the definition will give us a measure of the cost or time required to send the output of the message source to the destination. When applied to a channel, in the form information capacity of a channel, it will give a measure of how long it takes to transmit the message generated by one message source, or of how many message sources can be accommodated by one channel. We would like to be able to say that two comparable information sources generate twice as much information as one, and that two comparable transmission channels could transmit twice as much information as one.

The moment we identify information with the cost or the time which it takes to transmit a message from a message source to a destination, an interesting new fact emerges: Information is not so much a property of an individual message as it is a property of the whole experimental situation which produces the messages. For example, such utterances as: "How are you?," "Glad to meet you," "Happy Birthday," "Congratulations on the birth of your child," "Best Wishes to Mother on Mother's Day," carry very little information. These phrases belong to a very small set of polite stereotyped utterances, normally used in certain stereotyped circumstances. The telegraph company has taken advantage of this fact by listing on its telegraph blanks some 100 stereotyped messages for use in appropriate stereotyped situations. The customer chooses a message, and the signal transmitted by the telegraph company contains only the few symbols necessary to identify the particular message which has been chosen. At the receiving office, a clerk reconstitutes the stereotyped message for transmission to the destination. The fact that such a stereotyped message contains less information than most utterances containing the same number of words is reflected in the lower cost to send such a message.

In order to get an effective definition of information, then, we shall consider not only the message generated or transmitted, but also the set of all messages of which the one chosen is a member. The message source may be considered as an experimental setup capable of producing many different outcomes at different times or under different stimuli, and the messages as the outcome of one particular experiment.

Consider an experiment X whose outcome is to be transmitted (see Figure 2). We will be particularly interested in cases in which the outcome of experiment X is an honest message, say written English or a television picture, but for the moment consider experiments in general. First of all, suppose experiment X has n equally likely outcomes. In this special case the definition of information evolves naturally from the following argument.

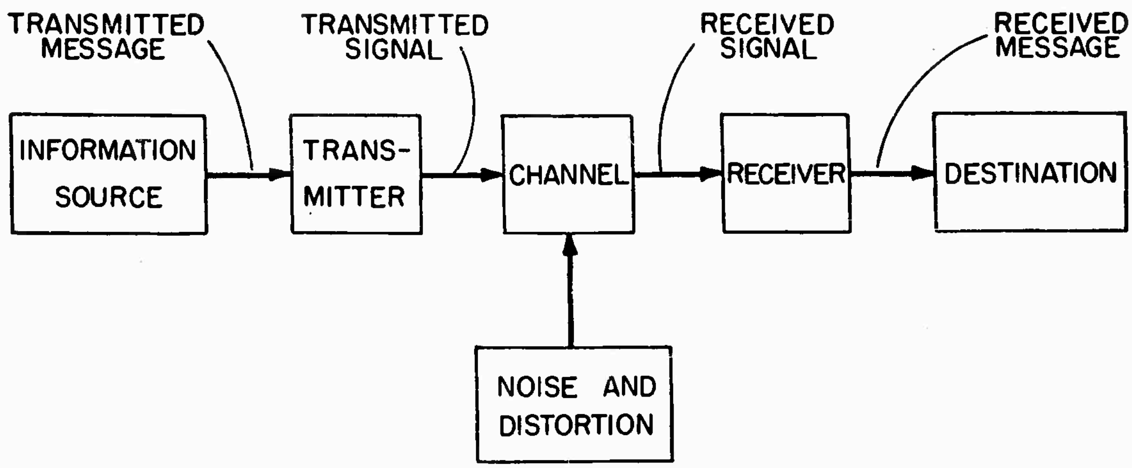


FIGURE 1 A GENERALIZED COMMUNICATION SYSTEM

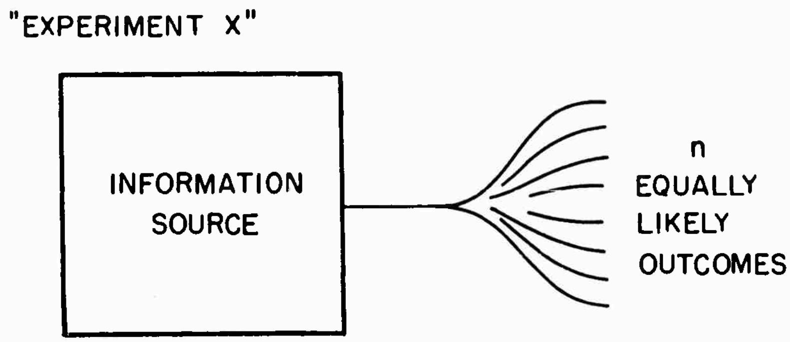


FIGURE 2 AN IDEALIZED INFORMATION SOURCE

The information in the message about X will be some function f(n). Suppose X is a compound experiment (see Figure 3) consisting of two independent experiments, Y and Z, which have n_1 and n_2 equally likely outcomes. The total number of outcomes of the compound experiment is the product of n_1 and n_2. Transmitting the outcome of X is equivalent to transmitting the outcomes of Y and Z separately. Thus the information of X must be the sum of the informations of Y and Z that is,

$$f(n) = f(n_1) + f(n_2)$$

where

$$n = n_1 n_2 .$$

This functional equation has many solutions. For example, f(n) might be the logarithm of n_1 or f(n) might be the number of factors into which n may be decomposed as a product of primes. However, there are other requirements of f(n). The time required to transmit the outcome of experiment X will certainly be an increasing function of n. Hence, we need consider only those solutions of the functional equation which are increasing functions of n. The only such solutions turn out to be constant multiples of log n, that is,

$$f(n) = c \log n .$$

The simplest possible experiment we can imagine is one which has two equally likely outcomes, like flipping a coin. We use the information associated with such an experiment as the unit for measurement of information and call it one bit. When this unit has been defined, the information in an experiment with n equally likely outcomes is then precisely $\log_2 n$ bits.

Let us now test this definition of information and see if it does the things that we expect from it. For example, what is the information associated with an experiment whose outcome is certain? The experiment might be, for example, to see whether the sun will rise between midnight and noon tomorrow. There is only one outcome possible:

$$n = 1.$$

The information associated with this experiment is

$$H = \log_2 1 = 0.$$

When the outcome of the experiment is a foregone conclusion, the information carried by the conclusion is zero.

What is the information associated with an experiment which has eight equally likely outcomes? According to our formula, the information should be equal to *

$$\log 8 = 3.$$

That is, it should have just three times as much information as that associated with flipping a coin. We can show that this is indeed the case by exhibiting the following code. Let the eight equally likely outcomes be identified as

HHH

HHT

HTH

THH

HTT

THT

TTH·

TTT

The form of the code makes it obvious that the outcome of this experiment can be associated uniquely with the outcome of a succession of three coin-flipping experiments, and conversely. From the point of view of transmitting the information, it makes no difference whether the code word represents the outcome of three coin-flipping experiments or of one experiment with eight equally likely outcomes. Therefore, the information contained in one experiment with eight equally likely outcomes is three times that contained in an experiment, like flipping a coin, with two equally likely outcomes, that is,

$$H = \log 8 = 3 = \log 2 + \log 2 + \log 2.$$

What happens if the various outcomes of the experiment are not equally likely? It is not immediately obvious that the definition of information can be extended. However, we can make a good try in the following way. Let us assume a situation (see Figure 4) where the experiment has n equally likely outcomes, grouped into two groups, an upper group of n_1 and a lower group of n_2, such that

$$n_1 + n_2 = n$$

---

*All logarithms are to the base 2 unless the contrary is specified.

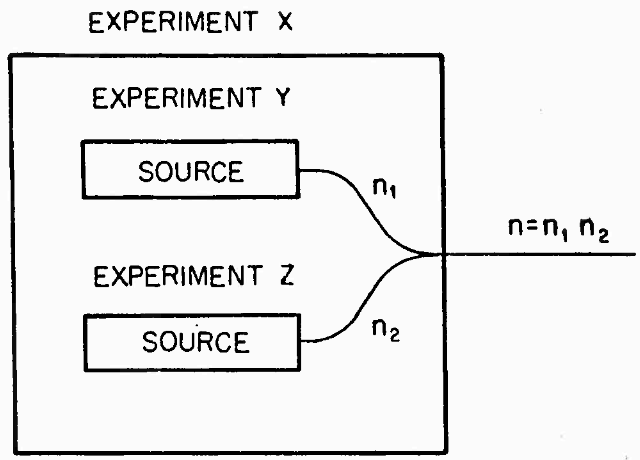


FIGURE 3 TWO INFORMATION SOURCES COMBINED INTO ONE

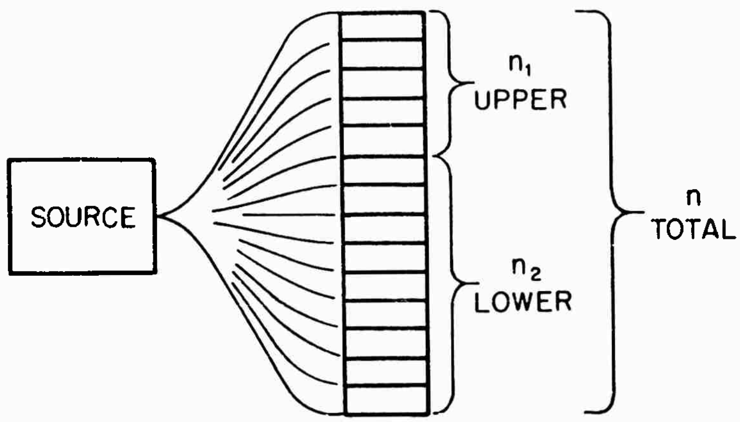


FIGURE 4 AN IDEALIZED SOURCE WITH OUTPUTS OF UNEQUAL PROBABILITY

Let us assume that we are not really interested in the particular message generated by the experiment, but only in whether the message is of the upper or of the lower group. We then have a situation where the significant output is one of two messages, having probabilities

$$p_1 = \frac{n_1}{n_1 + n_2}$$

for the upper message, and

$$p_2 = \frac{n_2}{n_1 + n_2}$$

for the lower, respectively. One way to find out how much information is associated with this is to start with the information associated with the n equally probable outcomes, and subtract the excess information with the n_1 or n_2 possible messages in the two sub-groups. The information associated with one message among n equally likely messages, is

$$\log n$$

The information associated with one message among n_1 equally likely messages is

$$\log n_1$$

This occurs not all the time, however, but only for a proportion of the time equal to n_1/n. The information associated with one of n_2 equally likely messages is

$$\log n_2$$

and this occurs for a proportion of the time equal to n_2/n.

Performing the arithmetic, we get

$$H = \log n - \frac{n_1}{n} \log n_1 - \frac{n_2}{n} \log n_2$$

$$= -p_1 \log p_1 - p_2 \log p_2.$$

Since p_1 and p_2 are less than unity, their logarithms are negative. Thus, one can see that the information H is positive.

This argument suggests a form for the amount of information in a message generated by experiment X having n possible outcomes which are not all equally likely. Let the various outcomes have probabilities $p_1, p_2, \ldots, p_n$. In this case, the amount of information in the message generated by the experiment X is defined to be

$$H(x) = -p_1 \log p_1 - p_2 \log p_2 - \ldots -p_n \log p_n$$

$$= \sum_{i=1}^{n} - p_i \log p_i$$

This sum bears a formal resemblance to a quantity called entropy in statistical mechanics. For this reason H(x) is also called the entropy function of $p_1, p_2, \ldots, p_n$.

Let us now look at this definition to see if we think it is appropriate as a measure of information. First of all, when the n outcomes are equally likely,

$$p_i = \frac{1}{n}$$

$$\log p_i = \log n$$

$$\sum_{i=1}^{n} -p_i \log p_i = \sum_{i=1}^{n} \frac{1}{n} \log n$$

$$= \log n$$

as it should.

It can be shown that for a fixed number of outcomes, the case of equally likely outcomes is the one in which H(x) attains its maximum value. This fits our intuitive notion very well: If all outcomes of the experiment are equally likely, the message gives a maximum of information; but if we have some a priori information that one outcome is more probable than another, then carrying out the experiment does not give quite so much information.

What if the experiment X consists of two independent experiments Y and Z? (See Figure 5.) Here the arithmetic is quite complicated, but ultimately one finds

$$H(x) = H(y) + H(z)$$

In words, the information associated with X is the sum of information of its constituent experiments Y and Z. If Y and Z are not statistically independent* (see Figure 6), then

$$H(x) < H(y) + H(z)$$

This again is reasonable. Some of the H(y) bits of information about the Y experiment give information about the possible outcome of the Z experiment and so are counted twice in the sum H(y) + H(z). So far, the definition of information which we have come up with seems satisfactory.

Let us recapitulate briefly. We started out with a model for a communication system which had an information source at one end and a destination at the other end. We have been looking for a definition of information which would be proportional to the time or the cost it takes to transmit the message from the message source to the destination. In order to get a firm hold on the problem, we successively restricted the information source until it was capable simply of putting forth n equally probable messages. In this case, we successfully defined information as log n. We have generalized this definition slightly to the entropy function, which defines the amount of information generated by a message source capable of generating one of a finite set of n messages with known probability distribution. We have verified that this definition of information fulfills some elementary intuitive notions of how a measure of quantity of information ought to behave.

In a way, it does not seem that we have gone very far. The message source that we considered is extremely restricted, for it allows nothing more general than signals made up of discrete, uniquely distinguishable characters, such as teletypewriter messages. It does not include any message represented by a continuous wave form, such as the sound pressure of speech or the video signal which will generate a television picture. But surprisingly, the major

---

*Imagine the experiments Y and Z performed many times, and suppose that the results of the Y experiment are classified into sets according to the outcome of the Z experiment. Examine the probability distribution of the results of the Y experiment in each set: if the distribution does not vary from set to set, Y and Z are statistically independent. In plain but less precise language, the expected result of Y is the same whatever the result of Z.

hurdle in defining quantity of information has already been passed. In spite of the fact that speech waves and television video signals are continuous signals, in any real life situation it is possible to distinguish only a finite number of tones or of picture intensities. The case of continuous messages can be reduced to the case of discrete messages already discussed above, and the definition of quantity of information can be directly adapted to this use.

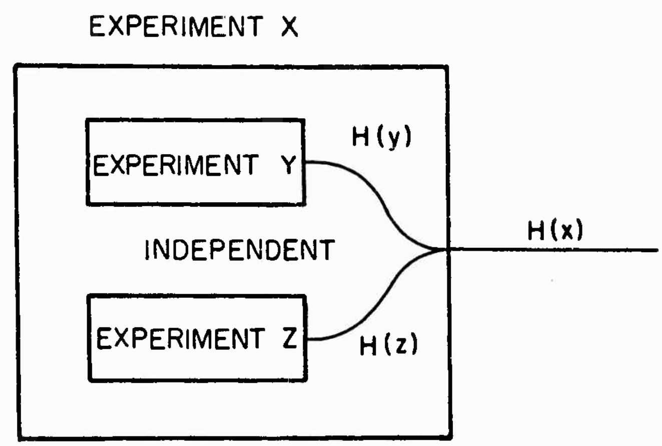


FIGURE 5 ILLUSTRATING THE SUMMING OF INFORMATION FROM TWO INDEPENDENT SOURCES: $H(x) = H(y) + H(z)$

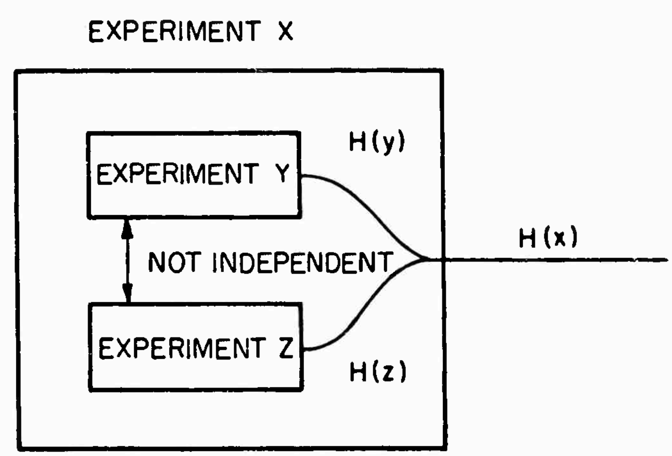


FIGURE 6 ILLUSTRATING THE SUMMING OF INFORMATION FROM TWO NONINDEPENDENT SOURCES: $H(x) < H(y) + H(z)$

IV. APPLICATIONS TO DISCRETE CHANNELS

Let us now look at some applications of the definition of information which has just been stated. Let us suppose that the experiment under consideration is that of shuffling a deck of 52 cards, and that the message is the particular order of the cards in the deck after shuffling. We shall define a perfect shuffle to mean that all of the possible orderings of the 52 cards are equally probable. Let us see how much information there is in a perfect shuffling experiment. The number of possible arrangements of the cards, according to well known formulas in combinatorial analysis, is 52!* The amount of information associated with this experiment is

$$\log 52! = 225.7 \text{ bits.}$$

Now let us look at another kind of shuffling experiment: Cut the deck into two packs, top (T) and bottom (B), at a random place, and then interleave T and B together. The interleaving operation consists of 52 steps, at each of which the bottom card of either T or B falls onto the top of the shuffled deck. The shuffle is completely described by a sequence of 52 letters T or B. (The i-th letter is T if at the i-th step the card fell from the bottom of packet T.) The position of the cut may be found from the sequence by counting the number of T's. There are only 2^{52} possible sequences of T and B, and hence only 2^{52} possible outcomes of the shuffling experiment. Even if we suppose all these outcomes to be equally probable, the maximum amount of information associated with this shuffling experiment is log of 2^{52}, or 52 bits.

Suppose we now ask the question, how many times do you have to cut and interleave a deck in order to achieve something approximating a perfect shuffle? We learned earlier that the information associated with a sequence of independent experiments is not greater than the sum of the informations developed by the experiments independently. Each cut and interleave shuffling operation generates at most 52 bits of information. A perfect shuffle generates 225.7 bits of information. Therefore, no sequence of fewer than 5 cutting and interleaving shuffles could possibly generate a perfect shuffle. We can say with confidence that to shuffle a deck fairly by cutting and interleaving, you must repeat the operation at least five times. There is no guarantee, of course, that this will produce a perfect shuffling operation: All we have found out is that if you cut and interleave fewer than five times, it certainly will not produce a perfect shuffle.

---

*$n! = n(n-1)\cdots 3.2.1$, e.g., $3! = 3.2.1 = 6$, $4! = 4.3.2.1 = 24$.

As another example, let us consider the information content of ordinary written English. To simplify the problems, let us talk about "telegraph English," which has no punctuation, no paragraphs, no lower case letters, and so forth. In this case, we have 27 symbols, the letters *a* to *z* and a space.

To get an upper limit to the amount of information, we can simply assume that all 27 symbols are equally probable. This sets an upper limit to the amount of information of $\log 27 = 4.76$ bits per letter.

This estimate is certainly pessimistic, because we know that the letters are not equally probable. By carrying out a count of letters in a sufficiently large sample of text, we can get an idea of the relative probabilities of spaces and letters in English text. Using this data, we can apply the formula we have developed to find out that the information in English text is not more than about four bits per letter.

This estimate can be refined somewhat with observations taken from cryptography. Consider the construction of a substitution cryptogram. In such a cryptogram, for each letter in the alphabet some other letter is substituted. The table which tells which letter is substituted for which is called the key, and it is not hard to find that the number of possible keys is $26!$. If we view the cryptogram (see Figure 7) as a compound experiment X whose two parts are Y, the communication of the clear text, and Z, the choice of a key from one of $26!$ possibilities, the total information associated with this compound experiment is no greater than $H(y) + \log 26!$ bits. We understand that substitution cryptograms of 40 letters can usually be solved, i.e., that given a 40-letter cryptogram, the information in both the text and the key can be recovered. Since 40 letters can contain no more than $40 \log 27$ bits of information, one concludes that

$$40 \log 27 \geq H(y) + \log 26!$$

and hence that the information in a 40-letter English message is

$$H(y) \leq 40 \log 27 - \log 26! \sim 100$$

The information in an English message is consequently no greater than 2.5 bits per letter.

By using more and more refined arguments, it has been shown* that the information content of ordinary English text is about one bit per letter.

---

*See Reference 6 in the Bibliography.

V. ENCODERS

It is useful here to introduce the idea of an encoder. An encoder may be described as a purely deterministic device which converts a message in one set of symbols into a new message, usually in a different set of symbols. For example, a handwritten English message may be converted into a pattern of holes punched on a tape, then into a sequence of electrical impulses on a teletype wire, back into English letters by a teletypewriter, and finally translated from English into French. The first three of these four operations are reversible encodings. That means that each incoming message can be encoded in only one way, and conversely, that no two different incoming messages are ever encoded alike. Translation from English into French, however, is not usually an encoding, because it involves random choices. For example, the English word "robbery" may be translated into either "vol" or "brigandage." Even assuming that all such choices were settled in advance, one would undoubtedly find some French words representing several English ones, for example, "vol" for both "robbery" and "theft." Then the encoding would not be reversible.

A reversible encoder transforms messages into encoded messages in a one-to-one way; one gets the same amount of information from the encoded message as from the original message. One would like to conclude that a reversible encoder driven by an information source is a new information source which generates information at the same rate as the driving source. However, this conclusion requires further assumptions about the encoder. For example, the encoder might just store the incoming message, and re-emit it at a slower rate. Such an encoder would ultimately require an unlimited amount of storage space. However, if a reversible encoder has only a finite number of internal states (for example, if it is made from a finite number of relays or magnetic cores or switching tubes with a finite memory), then the encoder output has the same information rate as its input.

We also need to talk about an idealized noiseless channel for transmission of discrete messages. An ideal channel has a finite list of symbols which it can transmit without error. A certain time is required to transmit each symbol. The times required to transmit the various symbols may not be the same.

The combination of a channel fed by a source may be regarded as a new source which generates the message at the receiving end (see Figure 8). The information rate of the received message will depend on the transmitting source. For example, suppose a channel can transmit English letters and word spaces at the rate of one symbol per second. When the channel transmits English text, it has a rate, as we have seen before, of about one bit per second. If the same channel is connected to a source which produces letters and spaces independently, with probability 1/27 for each kind of symbol, the rate is $\log 27 = 4.76$ bits per second. The largest rate at which one can signal over a channel, for all choices of the source, is called the capacity of the channel. The capacity of the English letter channel just discussed is 4.76 bits per second.

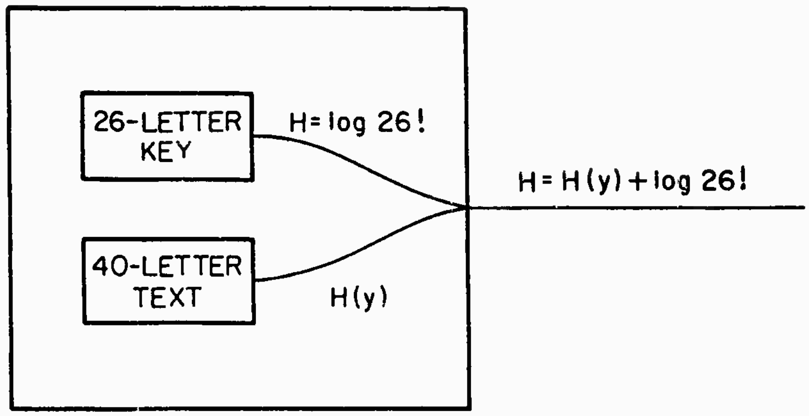


FIGURE 7 INFORMATION IN A 40-LETTER TEXT CODED WITH A SIMPLE SUBSTITUTION CODE

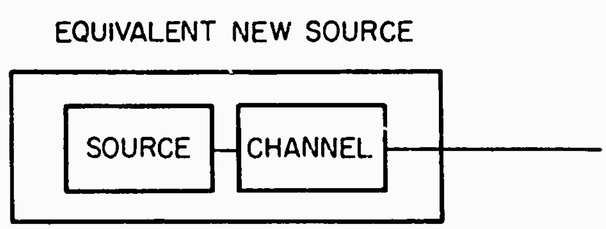


FIGURE 8 THE OUTPUT OF A COMMUNICATION CHANNEL REGARDED AS AN INFORMATION SOURCE

In the example, of the English text source connected to the English letter channel, one feels that much of the capability of the channel is wasted. With an English text source as input, the channel transmits information at a rate much lower than that attainable with other sources.

Is it possible to speed up the source and still use the same channel? The answer is yes, and an encoder provides the means for doing so. It is possible to encode English text reversibly in such a way that the encoded messages use fewer letters than the original messages. Then the encoded text may be transmitted at a higher information rate than the original text could.

In general, if we say that a channel has a capacity of C bits per second, we mean that the output of any source of information rate less than C bits per second may be transmitted over the channel by placing a suitable reversible encoder between the source and the channel. No reversible encoder will transform the output of any source having an information rate greater than C so that it can be transmitted through the channel without error.

To illustrate how the encoding process works, consider a very simple example. The source has two symbols: A, with probability 4/5; and B, with probability 1/5. Successive symbols are generated independently, at a rate of 80 per minute (see Figure 9).

The information rate of this source is

$$H = -0.2 \log 0.2 - 0.8 \log 0.8$$

$$= 0.72 \text{ bits per letter}$$

$$\frac{H}{T} = 0.72 \ \frac{80}{60}$$

$$= 0.96 \text{ bits per second}$$

So much for the source: now for the channel. The channel (see Figure 10) transmits two symbols, zero and one, without constraint, and requires precisely one second of transmission time to transmit either symbol. The channel capacity is thus one bit per second.

The simplest encoder we can imagine is the one shown in the following table:

Letters	Probability	Digits	Weighted Number of Digits
A	.8	0	.8
B	.2	1	.2
			1.0

The total weighted number of digits is

1.0 digits per letter = 80 digits per minute.

An example of a stream of letters and their encoding digits is

```
A B A A A A B A A A A A B A B A A A A A A A A A B A B A A A A
0 1 0 0 0 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
```

With such an encoder, 80 digits per minute are generated, and the channel will not tolerate them. A better encoder is shown in the next table.

Letters	Probability	Digits	Weighted Number of Digits
AA	.64	0	.64
AB	.16	10	.32
BA	.16	110	.48
BB	.04	111	.12
			1.56

Here, instead of encoding one message letter at a time, we group the message in bunches of two letters, and encode the two letters together. The relative probabilities of various groups of two letters vary over quite a range, as indicated in the second column. In order to gain efficiency in the coding, we use a short group of digits for a more common letter group, and reserve longer groups of digits for the less common letter groups. The last column, weighted number of digits, is the probability of a given digit-group multiplied by the number of digits in the group. Summing the last column over all letter groups, one finds an average digit-group length of 1.56 digits for two letters, or .78 digits per letter. The encoder turns out 62.4 digits per minute, still more than the channel will take. The same stream of letters is now encoded thus:

```
A B A A A A B A A A A A B A B A A A A A A A A A B A B A A A A
 10   0   0 110   0   0 110 110   0   0   0  10  10   0   0
```

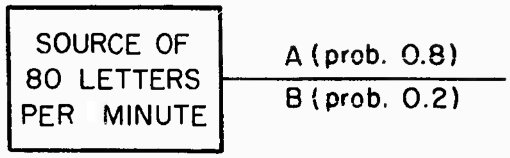


FIGURE 9 AN INFORMATION SOURCE

FIGURE 10 A COMMUNICATION CHANNEL
(Will the Information from the Source of Figure 9 pass through the Channel?)

The thirty letters are now encoded in 24 digits, 9 one's and 15 zero's. The reader can verify that if the digits are run together without spaces, they can still be separated unambiguously into symbols from our finite alphabet. Such a code is called <u>segmented</u>.

We can carry this a bit further, as shown in the next table.

Letters	Probability	Digits	Weighted Number of Digits
AAA	.512	0	.512
AAB	.128	100	.384
ABA	.128	101	.384
BAA	.128	110	.384
ABB	.032	11100	.160
BAB	.032	11101	.160
BBA	.032	11110	.160
BBB	.008	11111	.040
			2.184

In this example, each group of three letters is encoded in a single digit-group. The more common letter groups are encoded in short digit-groups, and the less common groups in longer digit-groups. Doing the arithmetic exactly as before, we find that the average digit-group length for three letters is 2.184 digits. This results in an average of .728 digits per letter, and the coder produces 58.24 digits per minute, which can be transmitted by the channel. We already know that the information content of this source is .72 bits per letter, and therefore, no reversible encoder could encode it in less than .72 digits per letter on the average. The encoder illustrated is only about 1% less efficient than the ideal. The stream of letters given before is now encoded thus:

```
A B A A A A B A A A A A B A B A A A A A A A A B A B A A A A
 101    0   110      0 11101     0      0   100   101     0
```

The stream of 30 letters is now encoded in 22 digits, 11 one's and 11 zero's. The fact that the number of one's and zero's grow closer and closer together is not an accident. We know that the maximum capacity of a two-symbol source is reached only when the two symbols have equal probability. Our coder must bow to this fact if it is to use the channel efficiently.

This encoder must have some storage capacity, and must introduce some delay. For example, three incoming letters must arrive and be stored before the outgoing digit-group is identified. Furthermore, the long digit-groups are transmitted more slowly than the incoming three-letter groups are generated; and signals must be stored until a string of AAA's allows the encoder and transmission channel to catch up. In this simple example, no finite storage capacity will guarantee flawless performance, but the probability of exceeding a storage requirement of a few hundred symbols is extremely small.

The above example illustrates the general coding theorem, which can be loosely expressed as follows: Given a channel and a message source which generates information at a rate less than the channel capacity, it is possible to devise an encoder which will allow the output of the message source, suitably encoded, to be transmitted through the channel.

VI. CHANNEL CAPACITY

A. CHANNEL CAPACITY OF AN ANALOG CHANNEL

In the coding theorem stated in the previous section, we have implicitly defined the channel capacity of a channel: If a channel can transmit C binary digits per second (but no more), its channel capacity is C. It is easy to apply this definition to a channel which transmits strings of zero's and one's at a fixed rate, as in the example above. It is equally easy to apply it to a teletypewriter transmission channel which transmits sequences of letters and spaces at a rate fixed by the terminal equipment. But this is not really very useful, because there has never been very much doubt about the capacity of such a channel. Suppose we have a more general channel: How do we determine its channel capacity?

This question really hinges on a determination of how many distinguishable signals the channel can transmit. To answer this question, we would like to have a way of identifying individual signals and distinguishing them one from another. What we really need is a catalog of signals.

Let us take as an example a channel capable of transmitting continuous waves with a finite bandwidth, free of distortion, but with uniform Gaussian noise of known power. Let us now identify and catalog the signals which can be transmitted through this channel.

We can get immediate help from the sampling theorem, a purely mathematical theorem now well known in the communication art, which will be stated here without proof (see Figure 11).

If a funtion of time f(t) contains no frequencies higher than W cycles per second, the function is uniquely determined by giving its ordinates a series of points spaced 1/(2W) seconds apart.

If we now let W be the bandwidth of the communication channel in question, we can identify any signal which the channel can transmit with a sequence of ordinates spaced 1/(2W) seconds apart. If we take a piece of this signal lasting only a finite time, say T, then the number of ordinates falling in this time range is 2TW.

We can now introduce some geometrical ideas to help us along with the cataloging process (Figure 12). A quantity which is identified by one number can be represented as a point on the straight line. A quantity identified by two numbers can be represented by a point on a plane: This is the familiar procedure used to plot graphs. A quantity identified by three numbers can be represented by a point in three-dimensional space. Similarly, our signal identified by 2TW

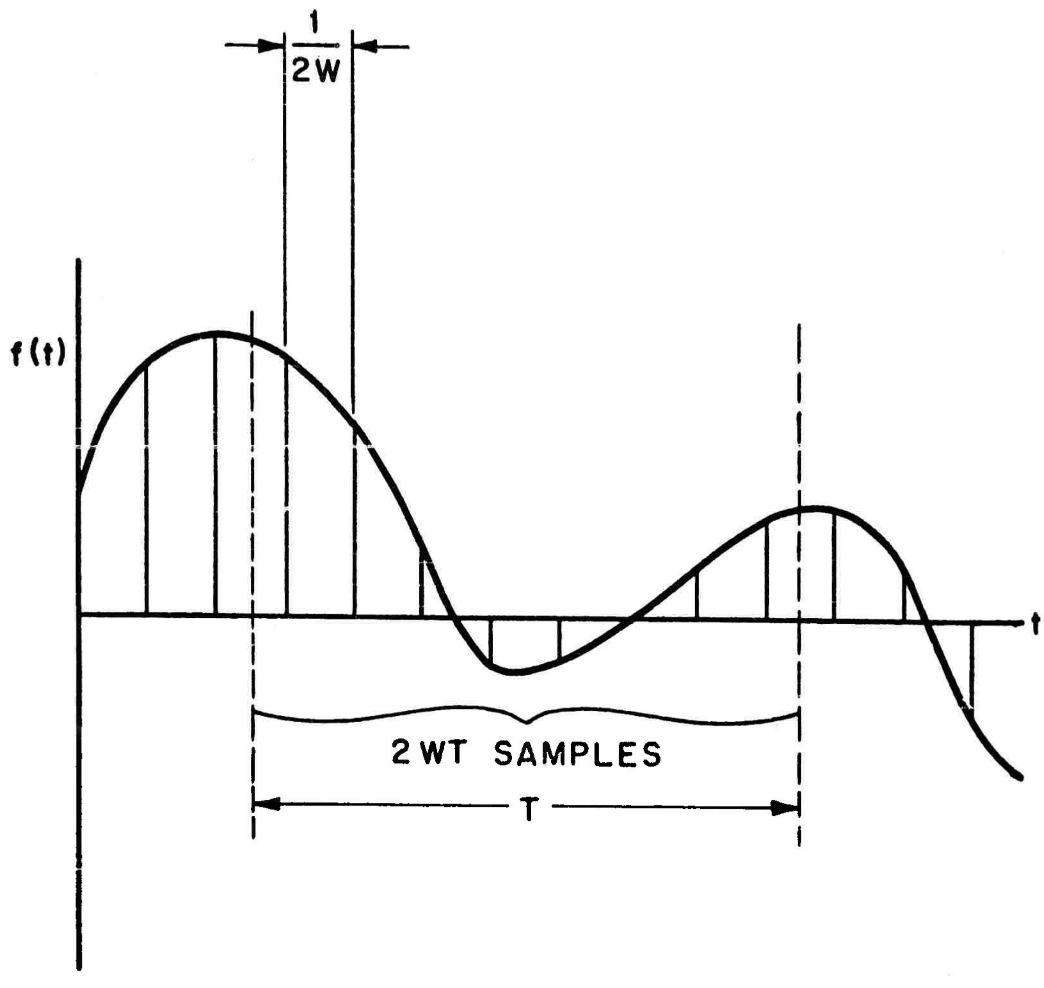
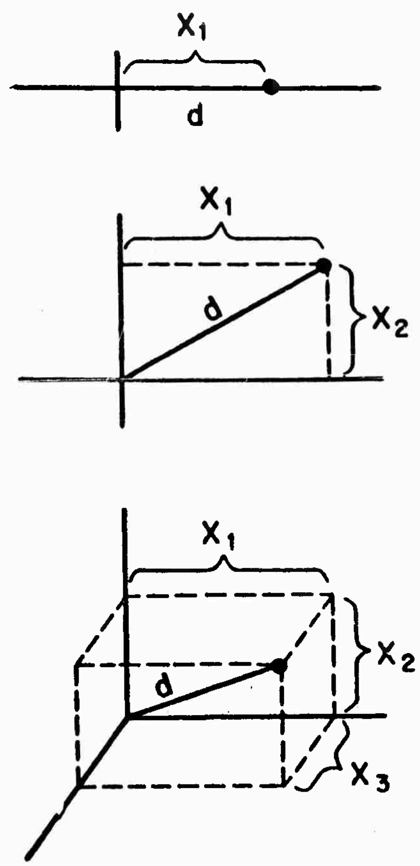


FIGURE 11 SAMPLING OF A BAND-LIMITED FUNCTION OF BANDWIDTH W



$$d = \sqrt{X_1^2}$$

$$d = \sqrt{X_1^2 + X_2^2}$$

$$d = \sqrt{X_1^2 + X_2^2 + X_3^2}$$

FIGURE 12 MULTI-DIMENSIONAL GEOMETRY

numbers can be identified with a point in a (necessarily imaginary) geometrical space of 2TW dimensions. We imagine the 2TW identifying numbers to be the coordinates of a point, measured along 2TW mutually perpendicular axes.

If we compute energy E in the signal, we find, except for a scale factor,* that

$$E = \frac{1}{2W} \sum x_n^2$$

where x_n is the nth coordinate, i.e., the nth sample of f(t). If we compute the distance from the origin to a point in the space which represents the same signal, we find

$$d = \sqrt{\sum x_n^2}$$

Thus

$$d^2 = 2WE$$

$$= 2WTP$$

where P is the signal power. In other words, in this geometric visualization of continuous signals, geometrical distance is proportional to the square root of the power. The distance between two points in space is proportional to the square root of the power of the difference of the two signals which the points represent. Signals of power less than P all lie within the sphere of radius $d = \sqrt{2WTP}$.

Now let us consider what happens to a signal as it goes through our channel. In Figure 13, we follow the geometric analogy, but represent the space of 2WT dimensions as two-dimensional space. A given input signal or output signal is represented by a point in the space. The distance between two points is proportional to the square root of the power of the difference of the two signals. Assume that the signal power is P, and that the power added by the noise in the channel is N. Assume that we know the position of the point in space representing the signal before it is transmitted through the channel. Where is this point at the output end of the channel? We do not know exactly, but we know approximately: It is somewhere in a sphere of radius $\sqrt{2WTN}$ centered around the point representing the transmitted signal. In the figure, this sphere is represented by a hatched circle. Just as the area of a circle is proportional to the square of its radius, and the volume of the sphere is proportional to the cube of its radius, so the

*If the signal is electrical and f(t) is the instantaneous amplitude in volts, the scale factor is the real part of the circuit admittance in mhos. All sums are over the range (1, 2TW), unless otherwise stated.

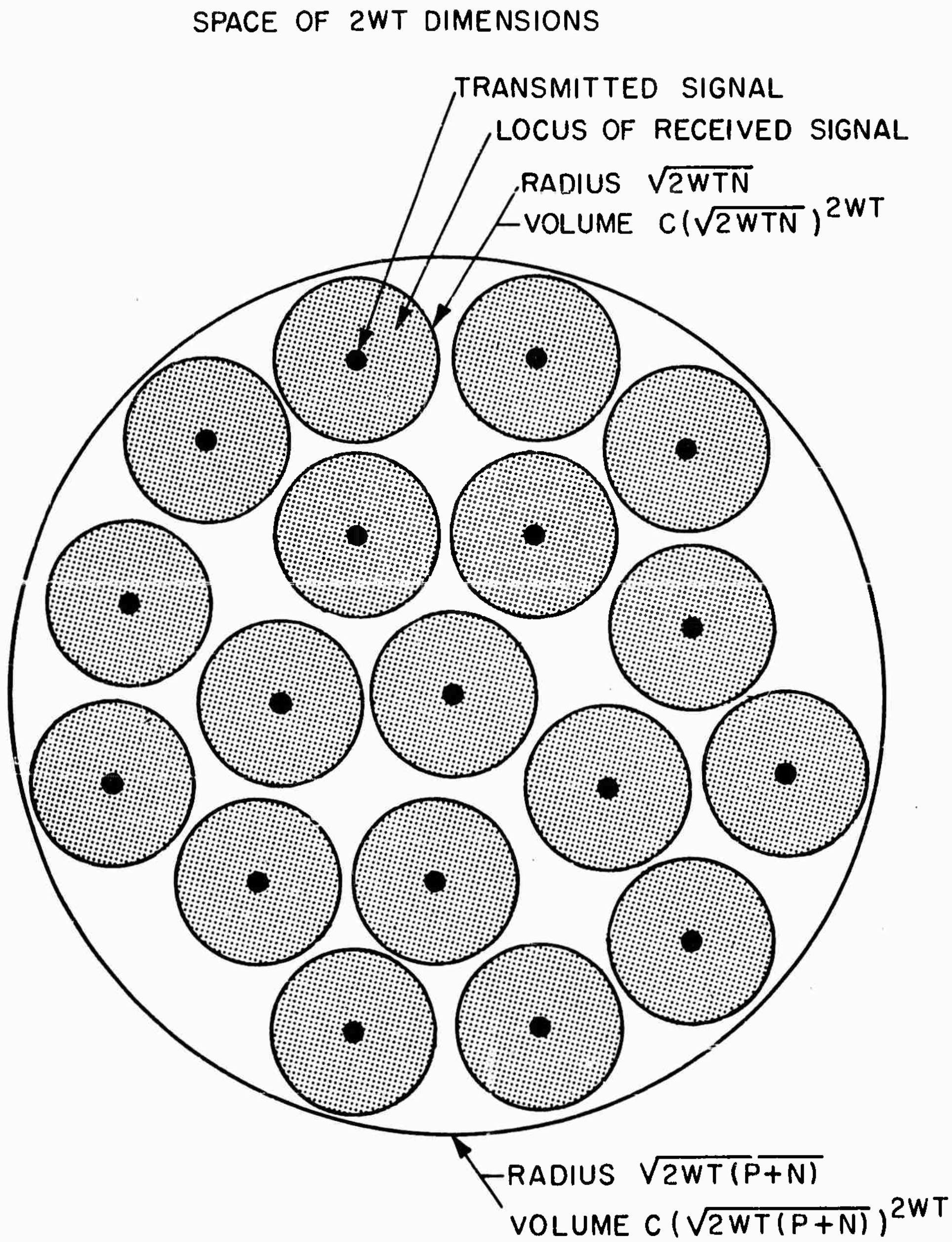


FIGURE 13 TRANSMITTED AND RECEIVED SIGNALS IN 2WT-DIMENSIONAL SIGNAL-SPACE

volume of this hypersphere is proportional to the 2WT power of its radius, say,

$$V = K\left(\sqrt{2WTN}\right)^{2WT}$$

where K is a constant whose numerical value is not important here.

The output of this channel consists of a signal plus noise, and has power approximately P + N. If we consider the whole family of possible outputs, they lie in a sphere of radius $\sqrt{2WT(P+N)}$. In the figure, this is represented by the large circle. The volume of this hypersphere is

$$V = K\left(\sqrt{2WT(P+N)}\right)^{2WT}$$

where K is the same unspecified constant as before. Now let us assume that we have a number, M, of transmitted signals such that the regions of uncertainty associated with them when they are perturbed by the noise are nonoverlapping. Then the large hypersphere contains M nonoverlapping small hyperspheres. The volume of the large hypersphere is at least M times the volume of one of the small hyperspheres. If we write down this inequality and solve for M we get

$$K\left(\sqrt{2WT(P+N)}\right)^{2WT} \geqslant MK\left(\sqrt{2WTN}\right)^{2WT}$$

$$M \leqslant \left(\sqrt{\frac{N+P}{N}}\right)^{2WT} = (1 + P/N)^{TW}$$

The ratio P/N is the familiar signal-to-noise ratio. We can find the average rate of information transfer thus:

$$\log M \leqslant TW \log (1 + P/N)$$

$$\frac{1}{T} \log M \leqslant W \log (1 + P/N)$$

This gives us an upper limit for the channel capacity of this channel.

To get a more useful result, we need a lower limit also. In fact, the lower limit turns out to be the same as the upper limit: We have an equality instead of an inequality. The details of the mathematical development are rather complex, and it is unnecessary to work them out here. However, we shall sketch the idea behind the proof, because it yields some important results.

The idea is as follows. One fixes a certain number, M, of points in this space as signals, without regard for spacing to avoid overlapping regions. A particular selection of M points constitutes a particular code for transmitting signals. After having picked M particular points, one computes the probability of error at the receiving end. This is the probability that a point in the space (observed at the receiving end of the channel) which is close to one code point is also close enough to another point so that it might be wrongly identified. The probability of error is then averaged over all possible choices of codes. After going through all the arithmetic, geometry, and trigonometry, we obtain the following result:

$$\frac{1}{T} \log M \geq W \log (1 + P/N) + \frac{1}{T} \log E_{av}$$

where E_{av} is the averaged probability of error. (Note that $E_{av} < 1$, so that $\log E_{av}$ is negative.)

We need to observe two things about this inequality. First, for some code choices, the error rate must be at least as low as the average error rate. Second, if we make T sufficiently large, we can make $1/T \log E_{av}$ as small as desired, and hence we can make

$$\frac{1}{T} \log M$$

as close as we desire to

$$W \log (1 + P/N)$$

and still make the average error rate as small as we please. Another way of saying this is

$$\text{l.u.b.} \left\{ \frac{1}{T} \log M \right\} = W \log (1 + P/N)$$

(where l.u.b. signifies least upper bound) for any value of average error rate, no matter how small.

We define this bound as the channel capacity, and can assert with confidence that there exist codes which permit transmission at a rate as close as desired to the channel capacity.

$$C = W \log (1 + P/N)$$

with an arbitraily small error rate.

Some secondary conclusions can be drawn from this argument. First, the points which represent the signals in the code must be fairly well distributed throughout the space. This means that the wave form of these signals will look more or less like noise, not like anything with systematic structure.

Secondly, in order to achieve high signalling rates and low error rates, it is necessary to use a space with a large number of dimensions and a large number of distinct signals. For a model whose performance is reasonably typical of what can be done, Fano* finds error rate and signalling alphabet size to be related by

$$P(e) \sim K\,2^{-\nu\alpha\, C/R}$$

where

P(e) is the probablility of error

K is a constant of the order of unity

ν is the number of binary digits constituting a message

or 2^{ν} is the number of distinct messages in the alphabet

C is the channel capacity

R is the actual signalling rate

and α is a particular function of R and C of the following form:

$$\alpha = \alpha\left(\frac{R}{C}\right) = \frac{1}{2} - \frac{R}{C} \qquad 0 \leqslant \frac{R}{C} \leqslant \frac{1}{4}$$

$$= \left[1 - \sqrt{\frac{R}{C}}\right]^2 \qquad \frac{1}{4} \leqslant \frac{R}{C} \leqslant 1$$

The model is a straightforward one in which the alphabet consists of two orthogonal signal wave-forms of equal energy, S, and equal duration, T; and in each time interval of length T, one and only one of the wave-forms is transmitted. For example, to achieve an error probability $P(e) = 10^{-5}$ with a signalling rate 95% of the capacity, i.e., R/C = .95, requires $\nu \sim 25,000$ bits per message.

---

*See Reference 4 in the Bibliography.

There is a third effect which is troublesome: In such a coding system, the threshold effect gets very sharp. As long as the noise power density is no greater than that assumed, the error rate is small, but if the noise power exceeds a certain level, then a point is reached very suddenly where the error rate jumps to a large value.

Nevertheless, this formula is very useful, for it provides a standard of comparison against which transmission channels and transmission systems can be judged. As we shall see presently, it also suggests ways to increase the channel capacity of certain practical types of communication systems.

B. CHANNEL CAPACITY OF SOME REPRESENTATIVE CHANNELS

Let us now compute the channel capacity of some typical transmission channels. First, what is the channel capacity of a 100-word-per-minute teletype (TTY) channel? This channel can transmit 600 letter or space characters per minute, or 10 characters per second. We saw before that the maximum information associated with one such character is 4.76 bits, so that the capacity of this channel is 47.6 bits per second - say 50 bits per second.

What is the channel capacity of an audio circuit for the transmission of speech? Being rather liberal, let us say that the signal-to-noise ratio P/N is 36 db, and that the bandwidth W is 4500 cycles per second. Such a channel is better than a telephone channel, and comparable to an AM broadcast radio channel. Working out the formula, we find that the channel capacity is 48, 000 bits per second - let us say 50, 000 bits per second.

What is the channel capacity of a channel used to transmit a video signal? Being rather liberal again, let us say that the signal-to-noise ratio P/N is 30 db, and that the bandwidth W is 5, 000, 000 cycles per second. Application of the formula in this case yields a channel capacity of 50, 000, 000 bits per second.

Thus, a voice circuit has about 1, 000 times the channel capacity of a teletypewriter channel, and a video circuit has about 1, 000 times the channel capacity of a voice circuit.

But is it possible to send the output of 1, 000 voice circuits through a single video channel, or to send the output of 1, 000 teletypewriter circuits through one voice channel? Not necessarily. As a matter of fact, many channels designed for video transmission will transmit very nearly 1, 000 voice circuits, but no one has ever squeezed 1, 000 teletypewriter channels into one voice channel of the kind just described and we do not expect that anyone ever will accomplish this feat.

We are usually satisfied to get 16 teletype channels into such a voice circuit, but sometimes use more elaborate equipment to get 48 circuits. By the use of extremely elaborate terminal equipment, we appear to be able to get 100 or even 200 teletype channels into such a voice circuit.

There are three reasons for this limitation. First, an actual voice transmission channel usually is not an ideal channel in the sense we have described it, uniform, invariant with time, with no perturbation other than random noise. Most radio and telephone voice channels have distortion and nonrandom noise, such as interference and cross talk, but of a nature which does not interfere with human voice communication. These perturbations may disturb other kinds of signals, and hence effectively reduce the channel capacity. Second, when we deal with discrete signals, we normally have a very small signalling alphabet, and at the same time demand low error rates. For example, if we send in the form of pulses through an apparatus that detects the pulses one at a time, so that $\nu = 1$, about five pulses are required for each character; and if we require a character error rate of less than 10^{-4}, then the error rate for an individual pulse must be $P(e) \leqslant 2.10^{-5}$. Solving the above equation for R/C gives $R/C \sim 0.03$; i.e., the number of teletype channels which could be multiplexed through one voice channel is about $0.03 \times 1000 = 30$. This value compares reasonably well with the observed value of 16, especially when we consider that the voice circuit for which the teletype multiplexer must be designed is usually a marginally satisfactory circuit having lower signal-to-noise ratio and smaller bandwith than the audio circuit described above. This consideration does not prevail in converting from television to voice and back, for the human listener does not decode the speech one bit at a time. He rather listens for whole phonemes, syllables, words, and even sentences before committing himself finally to a decision about what he has just heard.

Third, there is some loss, nevertheless, when a large channel is subdivided, just as wood is wasted when a tree is sawed into planks. However, in a system (such as the Bell System L-3 cable carrier transmission system) which is designed to carry voice or television signals, the trade-off is at the rate of 600 to 800 voice channels per television channel, and most of the remaining discrepancy is accounted for by "Guard bands," empty bands of frequency inserted between adjacent channels to make channel separation easier at the terminals.

Let us recapitulate briefly. We have defined quantity of information, and the rate at which information is generated by a discrete source. We have computed the information generated by certain kinds of sources. We have defined the channel capacity of a discrete channel. We have defined the channel capacity of a band-limited channel with Gaussian white noise, and used the definition to compute the channel capacity of certain kinds of channels. We have stated in loose form a theorem about encoding, to the effect that any channel can transmit the information from a source which generates information at a rate less than the channel capacity of the channel.

C. COMPARISON OF VARIOUS PRACTICAL COMMUNICATION CHANNELS

Let us now go back to the formula expressing the channel capacity of a band-limited noisy channel, and do some manipulation with it. For example, how much energy must be supplied to transmit one bit of information?

Let

P = signal energy in watts per cycle-per-second

W = signal bandwidth in cycles-per-second

Then

PW = signal power in watts

Since

C = channel capacity in bits per second

then

$\frac{PW}{C}$ = energy in joules per bit

Using the formula above for channel capacity C, one finds

$$\frac{PW}{C} = N \frac{P/N}{\log(1 + P/N)}$$

where

N = noise energy in watts per cycle-per-second.

In many practical situations, the noise energy per unit bandwidth is physically traceable to thermal effects, and is related to temperature by the formula

$$N = KT = 1.37 \cdot 10^{-23} \text{ T watts/cycle-per-second}$$

where K is Boltzmann's constant and T is the absolute temperature. This relation leads to the definition of an effective temperature or noise temperature

$$T_e = N/K$$

even when the actual noise N may not be of thermal origin.

The number of joules required to transmit one bit is directly proportional to the noisiness or noise temperature of the channel, a relation which is quite understandable, and also to a certain function of the signal-to-noise ratio P/N. This function is plotted (Figure 14) as a function of the signal-to-noise ratio for easier analysis of its behavior. It is a steadily increasing function of P/N. Its minimum value is 0.693, which is approached when P/N is zero, that is, when the signal is very small compared to the noise. When the signal power density is as great as the noise power density, that is, when P/N equals one, the value of this function has risen from .693 to unity. Beyond that point it rises very rapidly. For the signal-to-noise ratios that we like to think of in communications, 30 or 40 db, this function exceeds 100. The energy required to transmit one bit of information is 100 times greater when the signal-to-noise ratio is 30 db than when the signal-to-noise ratio is less than zero db.

This observation is not new, but it still comes as a shock to a great many people. Many will insist that it is not in accordance with experience. Why do we persist in using communication systems which use so much more energy than necessary to transmit information?

There are three principal technical reasons why most communication systems do not approach this ideal.

First, the modulation system does not make efficient use of bandwidth in reducing power required.

Second, the signal in its original form does not make efficient use of the channel provided, that is, the signal characteristics and the channel characteristics are not well matched.

Third, the information content of the signal is not commensurate with its characteristics. Most signals which it is desired to transmit contain a great deal of unnecessary detail, that is, they are greatly redundant. Redundancy may be useful, since it adds to the reliability, or accuracy of the message, but it is not usually present in a very efficient form.

All of these technical objections could be overcome or alleviated, at least in some degree, but the ultimate decision faced by the communications system engineer is based not on the desire to transmit a bit with the least possible amount of energy, but on the desire to satisfy a particular communication need at the minimum cost. In most communication systems designed in the past, the cost of power has not been one of the principal system costs. However, when power does become an important part of the cost of the communication system, the designers will be driven to systems which operate with broader bandwidth and lower signal-to-noise ratio, in order to make the best possible use of power.

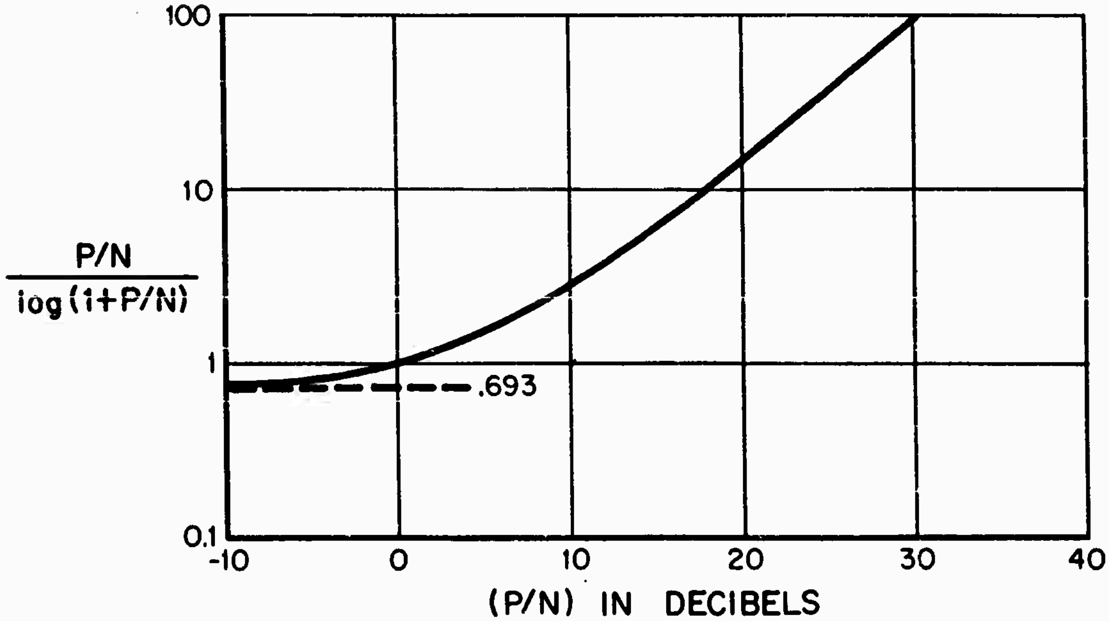


FIGURE 14 NORMALIZED ENERGY PER BIT REQUIRED TO SIGNAL OVER A NOISY CHANNEL

In electronics systems involving the use of unattended equipment in satellites, power becomes an important factor because it must be generated by solar batteries or by some other relatively uneconomical means--uneconomical not only because of initial cost, but also because the power supply may take up a significant part of the total available space and weight. In passive communication satellite experiments such as Project ECHO, power is once again one of the limiting factors in performance. There is good reason to believe, therefore, that designers of communication equipment for use in active and in passive satellite communication relay systems will try to exploit the advantages of broad bandwidth, low signal-to-noise-ratio communication in the future.

In sending signals by radio, we can use various systems of modulation. These require various bandwidths and powers, and have various advantages depending upon the signal characteristics and system requirements. Let us see how close they approach the ideal of using only 0.693N joules to send a bit.

We will consider first three comparatively well-known modulation schemes: single sideband modulation (SSB), frequency modulation (FM), and frequency modulation with feedback (FMFB).

In single sideband modulation (SSB), a constant radio frequency is added to all frequencies in the baseband (voice, TV, or other) signal. For example, a baseband signal $a \cos 2\pi ft$ might be represented as a modulation wave $a \cos 2\pi(f_0 + f)t$, where f_0 is the carrier frequency. Figure 15 a and d illustrates the spectra of such signals. The rf bandwidth required is the same as the baseband bandwidth b. The signal-to-noise ratio in the recovered baseband signal is the same as the rf signal-to-noise ratio (assuming that no noise is added in amplification). That is,

$$\frac{S}{N} = \frac{P}{N}$$

where

S = baseband signal spectrum power density in watts per cycle-per-second (joules)

and P and N are defined as before. Thus

$$C = W \log (1 + S/N)$$

$$= W \log (1 + P/N)$$

and

$$\frac{PW}{C} = N\frac{P/N}{\log(1+P/N)}$$

$$= (0.693N)\left[1.44\,\frac{P/N}{\log(1+P/N)}\right]$$

The system is less efficient than the ideal by a factor

$$1.44\,\frac{P/N}{\log(1+P/N)}$$

For output signal-to-noise ratios required for good quality speech or television, this factor makes the system several hundred times less efficient than the ideal. The main advantage of SSB is its economy of bandwidth.

In amplitude modulation (AM), the baseband signal $a \cos 2\pi ft$ is represented by the modulated signal $(1+\cos a 2\pi ft)(\cos 2\pi f_0 t)$. By trigonometric identities this signal can be shown to be equal to

$$(a/2)\cos 2\pi(f_o - f)t + \cos 2\pi f_o t + (a/2)\cos 2\pi(f_o + f)t$$

The AM spectrum is illustrated in Figure 15b. The constant carrier term $\cos 2\pi f_0 t$ can be removed by filtering to get a suppressed carrier AM signal, whose spectrum is illustrated in Figure 15c.

In AM, an rf band twice as big as the base bandwidth is required, because two sidebands are transmitted. At full modulation, AM requires three times as much power, and with ordinary signal statistics, many times as much power, as SSB. However, when the carrier is suppressed, the system has the same power requirement as SSB, but still requires twice the bandwidth. The chief advantage of AM over SSB is the circuit simplicity.

In frequency modulation, the baseband signal $a \cos 2\pi ft$ is represented by the modulated signal

$$\cos(2\pi f_o t + M\cos 2\pi ft)$$

This cannot be expressed as a finite number of cosinusoids. However, it can be expressed as

$$\sum_{n=-\infty}^{\infty} J_n(M)\cos\left[2\pi(f_o + n f)t\right]$$

where $J_n(M)$ is the Bessel function of order n and argument M.

This is illustrated in 15e for $M = 2$. Now it is a mathematically valid and practically justifiable observation that when $|n| > M + 1$, $J_n(M)$ is very small, and we can ignore those components. This results in a practical estimate of rf bandwidth.

$$B = 2\,(M + 1)b$$

Another way of justifying this heuristically is to say that the instantaneous carrier frequency varies from $f_0 - Mf$ to $f_0 + Mf$ and carries with it a local sideband pattern of width 2b, just as an AM signal does. The estimate is rough, but is amply justified by its practical usefulness and validity,

$$B = 2(\Delta f + b) = 2(M + 1)b$$

If (P/N) is the rf carrier-to-noise power ratio, the baseband signal-to-noise ratio (S/N) is

$$\frac{S}{N} = 3(P/N)M^2\,(M + 1)$$

This formula looks abstruse, and is somewhat difficult to derive, but it is really quite plausible, as can be seen from the following argument. Suppose we imagine a system in which the total transmitted power (carrier power) is fixed, but the modulation index M is variable. The output of the detector is a measure of the frequency deviation of the carrier, and its amplitude is therefore proportional to M. The signal power S therefore varies as M^2:

$$S \propto M^2$$

On the other hand, the spectral power density P of the transmitted signal is related to the carrier power P_c by

$$P = \frac{P_c}{B} = \frac{P_c}{2(M + 1)b} \times \frac{1}{M + 1}$$

Hence

$$\frac{S}{P} \propto M^2\,(M + 1)$$

or

$$\frac{S}{N} \propto \frac{P}{N}\,M^2(M + 1)$$

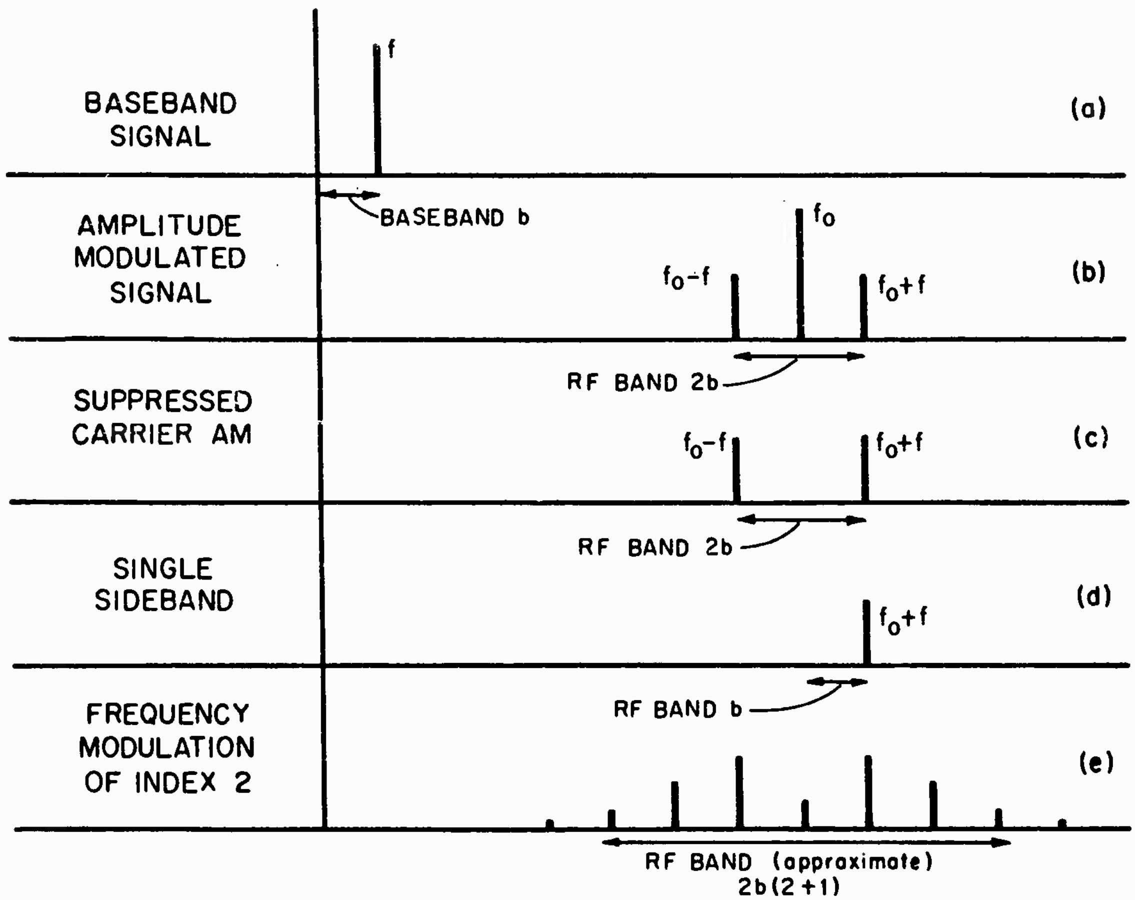


FIGURE 15 SPECTRUM OF AM, SUPPRESSED CARRIER, SSB, AND FM WAVES WHEN THE BASEBAND SIGNAL IS A SINGLE COSINUSOID

There only remains the evaluation of the constant of proportionality. A more detailed analysis shows that the correct value is three. The analysis is complicated by the fact that the demodulated noise spectrum density of an FM channel is not uniform, but is proportional to demodulated frequency.

For an FM detector system to work, it is necessary that the carrier amplitude be large in comparison with the noise amplitude. It is not hard to see why: the discriminator must be able to follow unambiguously the coherent pattern of peaks and dips in the sinusoidally oscillating signal. If the noise is too big, a loop of the sinusoid will be cancelled out from time to time, or an extra peak or dip added. Under these conditions, the discriminator will make an erroneous identification of phase and will skip or add an apparent full cycle. Practically speaking, this hazard is reduced to negligible proportions only if the carrier-to-noise ratio is at least

$$\frac{P}{N} = 16, \text{ or } 12 \text{ db}$$

As the index M is increased, the required rf bandwidth is increased; hence, the total rf noise is increased, and the minimum permissible transmitted power is increased. On the other hand, increasing the deviation makes the baseband signal-to-noise ratio greater than the carrier-to-noise ratio.

The channel capacity at minimum power level is

$$\begin{aligned} C &= b \log (1 + S/N) \\ &= b \log \left[1 + 48M^2 (1 + M) \right] \end{aligned}$$

Hence, the energy per bit is

$$\frac{PB}{C} = (0.693N) \frac{46 (1 + M)}{\log \left[1 + 48M^2 (1 + M)\right]}$$

The energy is greater than the ideal of 0.693N by a factor

$$\frac{46 (1 + M)}{\log \left[1 + 48M^2 (1 + M)\right]}$$

This factor has an optimum value of about 15, consistent with an index, M, of two and an output, S/N, of 600 or 27 db. Thus, ordinary FM is at best about 15 times less efficient in the use of power than the ideal. The efficiency of FM is relatively

insensitive to variation of index M from 1 to 4. The corresponding range of signal-to-noise ratios is 20 to 35 db. This range is of considerable practical interest for voice and many other analog signals.

Figure 16 shows a block diagram of frequency modulation with feedback, called also Chaffee system or FMFB.

In an FMFB system, we use the output of the discriminator to cause a beating oscillator partially to track changes in carrier frequency. Of course, it cannot track perfectly, for in that case the output of the mixer would have constant frequency and there would be no signal for the discriminator to detect. However, if a frequency change δf at the detector causes a change $\mu\delta f$ in the voltage tuned oscillator, then the deviation M_i in the intermediate frequency amplifier is reduced to

$$M_i = \frac{M}{1 + \mu}$$

Here μ is completely analogous to the gain in the feedback loop of a linear amplifier, and the amount of feedback in db is

$$\text{feedback} = 20 \log_{10} \mu \text{ db}$$

Thus we can cut down the intermediate frequency bandwidth B_i to a value

$$B_i = 2\left(\frac{M}{1 + \mu} + 1\right) b$$

Inasmuch as the IF bandwidth is less than the total rf bandwidth, the noise in the IF band is less than that in the rf band. We will still need a 12-db carrier-to-noise ratio at the discriminator, but the rf carrier-to-noise ratio can be less by the ratio of the IF bandwidth to the rf bandwidth.

Another way of expressing this idea is illustrated in Figure 17. The spectrum of the FM wave, as described before, extends from $f_o - 3_f$ to $f_o + 3_f$. This spectrum is illustrated in Figure 17a. However, over a short period of time an investigation of spectral energy density will show the energy to be concentrated about the instantaneous frequency in a band of breadth about 2b. This is depicted in Figure 17b. A filter of bandwidth 2b located at the right center-frequency would pass almost all the signal energy. The effect of the feedback loop in the detector is to shift the effective center frequency of the IF filter almost in synchronism with the instantaneous frequency of the incoming carrier.

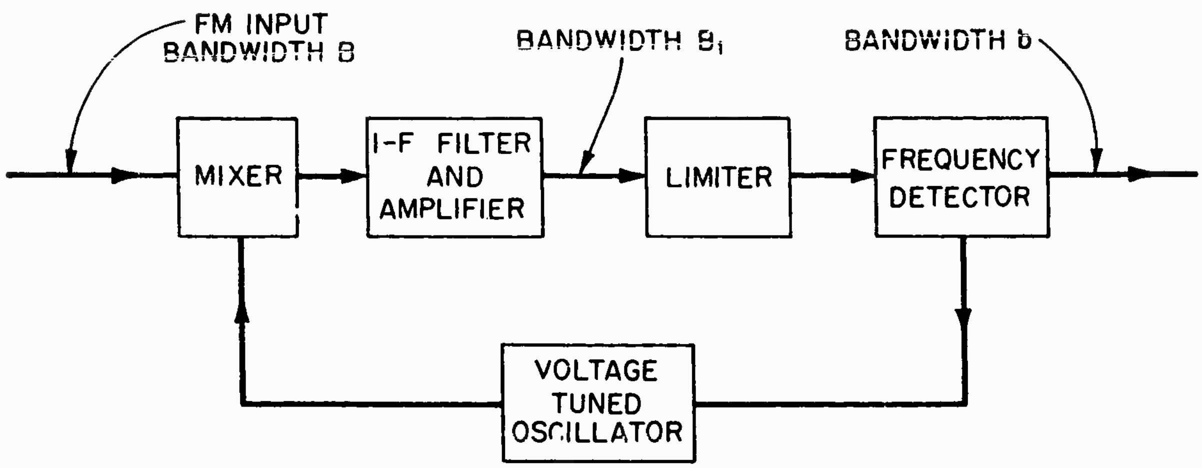


FIGURE 16 FREQUENCY-MODULATION-WITH-FEEDBACK (FMFB): BLOCK DIAGRAM OF A DETECTOR

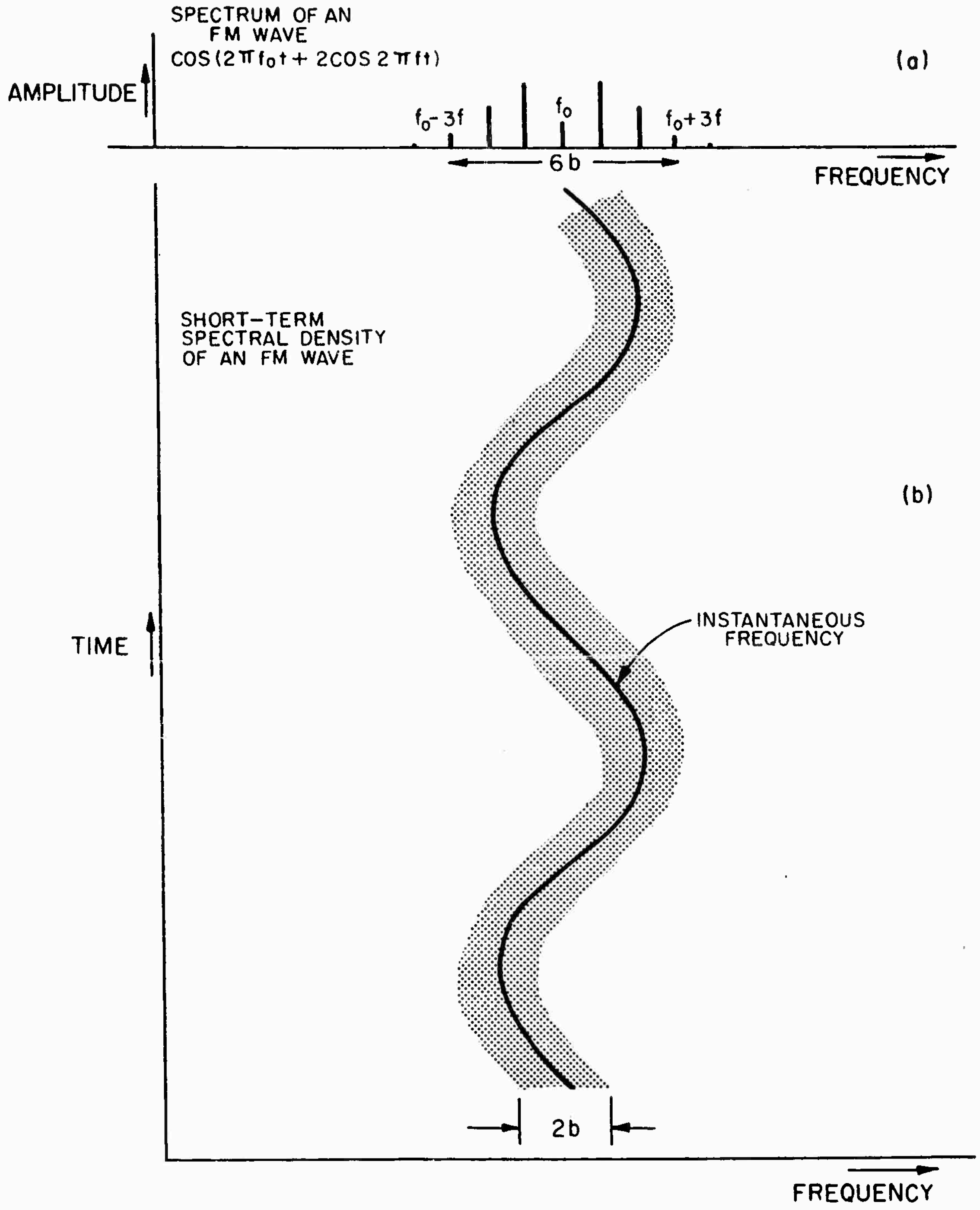


FIGURE 17 SPECTRUM AND SHORT-TIME SPECTRAL DENSITY OF AN FM WAVE

The minimum allowable signal-to-noise ratio now becomes

$$\frac{PB}{NB_i} = 16$$

An analysis like the one performed above leads to a required energy-per-bit of

$$\frac{PB}{C} = 0.693 \frac{46\left(\frac{M}{1+\mu} + 1\right)}{\log\left[1 + 48M^2\left(1 + \frac{M}{1+\mu}\right)\right]}$$

This energy is greater than the ideal by a factor

$$\frac{46\left(\frac{M}{1+\mu} + 1\right)}{\log\left[1 + 48M^2\left(1 + \frac{M}{1+\mu}\right)\right]}$$

This expression is only approximate, because when M is very large, the minimum allowable discriminator signal-to-noise ratio is greater than 12 db. When this is accounted for, this factor is found to go asymptotically to a theoretical value of two as M is increased. Experimentally, it appears that one can achieve a value around three, i.e., that one can operate with only three times the minimum theoretical power requirement given by information theory.

That is, it is possible to receive information with a receiver power of:

$$P = 3(0.693)CN$$

$$= 3(0.695)CKT_e \text{ watts}$$

where T_e is the effective noise temperature, and K is Boltzmann's constant.

Phase lock reception is similar to the foregoing system except that the local oscillator is in effect made to track the received signal in phase.

Some pulse transmission systems, such as pulse position modulation, appear to be capable of as great a power efficiency as FMFB. Whether or not they are competitive will depend upon equipment economy and, in some cases, upon the kind of information that is to be transmitted.

It should be noted that the channel capacities attributed to various modulation systems above are not binary digit signalling rates. We have accepted at face value the value which the channel capacity formula gives for the demodulated baseband channel, and compared that with the rf power. This comparison is still fair, however, if we are dealing exclusively with analog channels.

VII. A NOTE ON PROBABILITY DISTRIBUTION

In dealing with collections of numbers having properties of randomness, such as observations of electrical noise, it is convenient to introduce certain concepts from statistical analysis. In particular, let us assume we have a collection of numbers x_1, x_2, x_3, ... x_N, and define the following:

$$m = \text{the mean} = \frac{1}{N} \sum_{m=1}^{N} x_m$$

$$s^2 = \text{the variance} = \frac{1}{N} \sum_{m=1}^{N} x_m^2 - m^2$$

The mean is what we call in plain language the average. The variance is more esoteric: the square root of the variance, s, is called the standard deviation, and is a measure of the extent to which the numbers x_N, scatter from the mean value m.

Under many circumstances the set of N numbers is taken from a much larger or infinite set, called the population. This set of N numbers is then called a sample. The population mean μ and population variance σ^2 are defined just as the sample mean m and variance s^2. If necessary, limiting operations are used. If the number of elements, N, in the sample is large, we are often justified in treating the sample mean m and variance s^2 as about equal to the population mean μ and variance σ^2.

If each element x_m of the population is the sum of a large number of statistically independent numbers then (with certain technical restrictions) the distribution of values of the elements x_m will approach a particular distribution, called the Gaussian or normal distribution, characterized thus: in any random sample of N elements, the number of elements having a value between x_o and $x_o + \Delta x$ is approximately

$$N \; P\left[(x_o - \mu)/\sigma\right] \Delta x/\sigma$$

where P(u) is the normal probability distribution function

$$P(u) = \frac{1}{\sqrt{2\pi}} e^{-\mu^2/2}$$

The normal probability distribution has been extensively studied, and is a satisfactory model for a wide variety of statistical phenomena. Sums and differences of normally distributed independent numbers are also normally distributed. For example, we can take sums of the elements x_n M at a time, thus

$$y_o = \sum_{1}^{M} x_n, \quad y_1 = \sum_{M+1}^{2M} x_n, \quad y_k = \sum_{kM+1}^{(k+1)M} x_n$$

Then the population of all possible values of y_k has a mean $M\mu$ and a variance $M\sigma^2$. This and other properties of normal distributions will be referred to often in the next sections, and are described and proved in texts on probability.

VIII. DETECTION AS A COMMUNICATION PROCESS

Detection of a signal such as a radar echo in a background of noise may be treated as a communication process also. Suppose, for example, a situation exists where a signal s(t) may or may not be present in a background of noise n(t). Let us suppose for illustration that the noise is Gaussian with a uniform power density spectrum N up to a maximum frequency W, that the signal falls in the same frequency range, and that our observation is limited to the period of time $0 \leq t \leq T$, which is supposed to include all of the nonzero part of the signal s(t).

Using the sampling theorem as before, we can represent the signal by a point in 2WT-dimensional space. It is convenient to make a slight scale-change and represent a function f(t) by*

$$f(t) = \sum_{k=1}^{2TW} f_k \varphi_k(t)$$

where

$$\varphi_k(t) = \sqrt{2W} \frac{\sin 2\pi W(t - k/2W)}{2\pi W(t - k/2W)}$$

$$f_k = \frac{1}{\sqrt{2W}} f(k/2W)$$

Figures 18 and 19 show graphically how a function f(t) is built up of such elements φ.

It is not hard to show that

$$\int_{-\infty}^{\infty} \varphi_k(t)\varphi_\ell(t)\, dt = \begin{cases} 0 \text{ if } k \neq \ell \\ 1 \text{ if } k = \ell \end{cases}$$

Given two functions f(t) and g(t), we can define a scalar product

$$f(t) \cdot g(t) = \sum_{1}^{2TW} f_k g_k$$

*Unless otherwise indicated all sums are over the range (1, 2TW) and all integrals over the range $(-\infty, \infty)$.

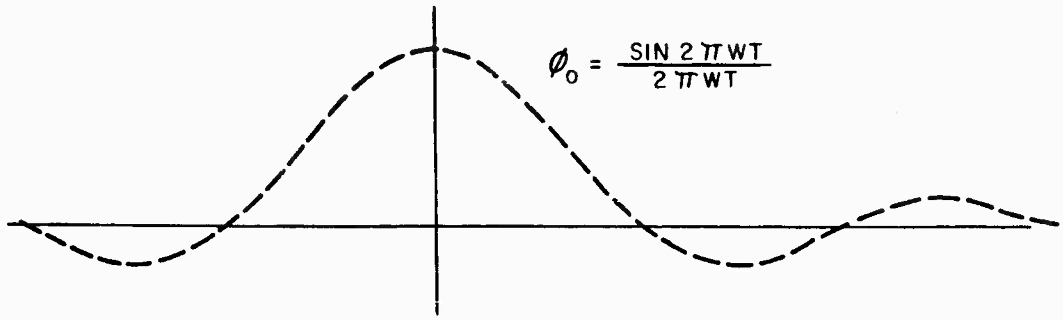


FIGURE 18 A PULSE FOR CONSTRUCTING BAND-LIMITED FUNCTION FROM EQUALLY SPACED SAMPLES

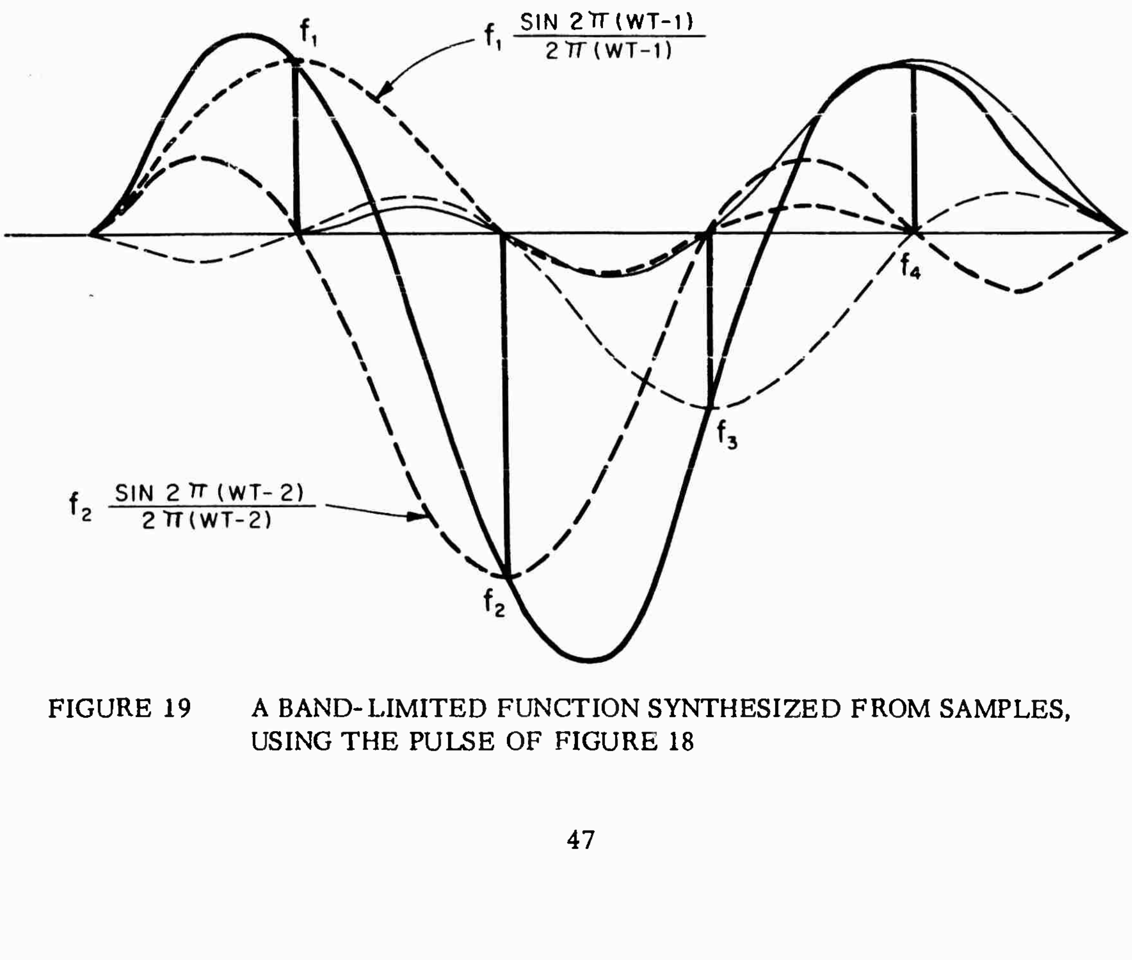


FIGURE 19 A BAND-LIMITED FUNCTION SYNTHESIZED FROM SAMPLES, USING THE PULSE OF FIGURE 18

From the above integral relation it follows that

$$\int_{-\infty}^{\infty} f(t)\, g(t)\, dt = \int_{-\infty}^{\infty} \sum f_k \varphi_k(t) \sum g_\ell \varphi_\ell(t)\, dt = \sum_{1}^{2TW} f_k g_k$$

which provides an alternative formula for the scalar product. Following this notation we let

$$s(t) = \sum s_k \varphi_k \,, \quad n(t) = \sum n_k \varphi_k$$

We may call the total signal energy S, and we see that, in suitable units,

$$\int s^2(t)\, dt = S = \sum s_k^2$$

The total noise energy is the product of noise spectral density, bandwidth, and time.

$$NWT = \int n^2(t)\, dt = \sum n_k^2$$

The expected value of n_k^2 for any k is therefore N/2. To avoid a sticky problem, we can assume the noise sample amplitudes n_k have expected value zero and variance N/2 and that they are independent and normally distributed. This is a satisfactory definition of white Gaussian noise of power density spectrum N and bandwidth W.

Now let us consider the detection problem where the noise field n(t) is present, and the signal s(t) may or may not be present. We observe a received signal f(t) where

$$f(t) = s(t) + n(t) = \sum (s_k + n_k)\, \varphi_k \quad \text{when the signal is present.}$$

$$= n(t) \qquad = \sum n_k \varphi_k \qquad \text{when the signal is absent.}$$

Figure 20, a and b, illustrates a pair of such wave-forms. When no signal is present, the expected value of each coordinate f_k is zero, and its variance is N/2. When the signal is present, the expected value of f_k is s_k, and the variance is still N/2.

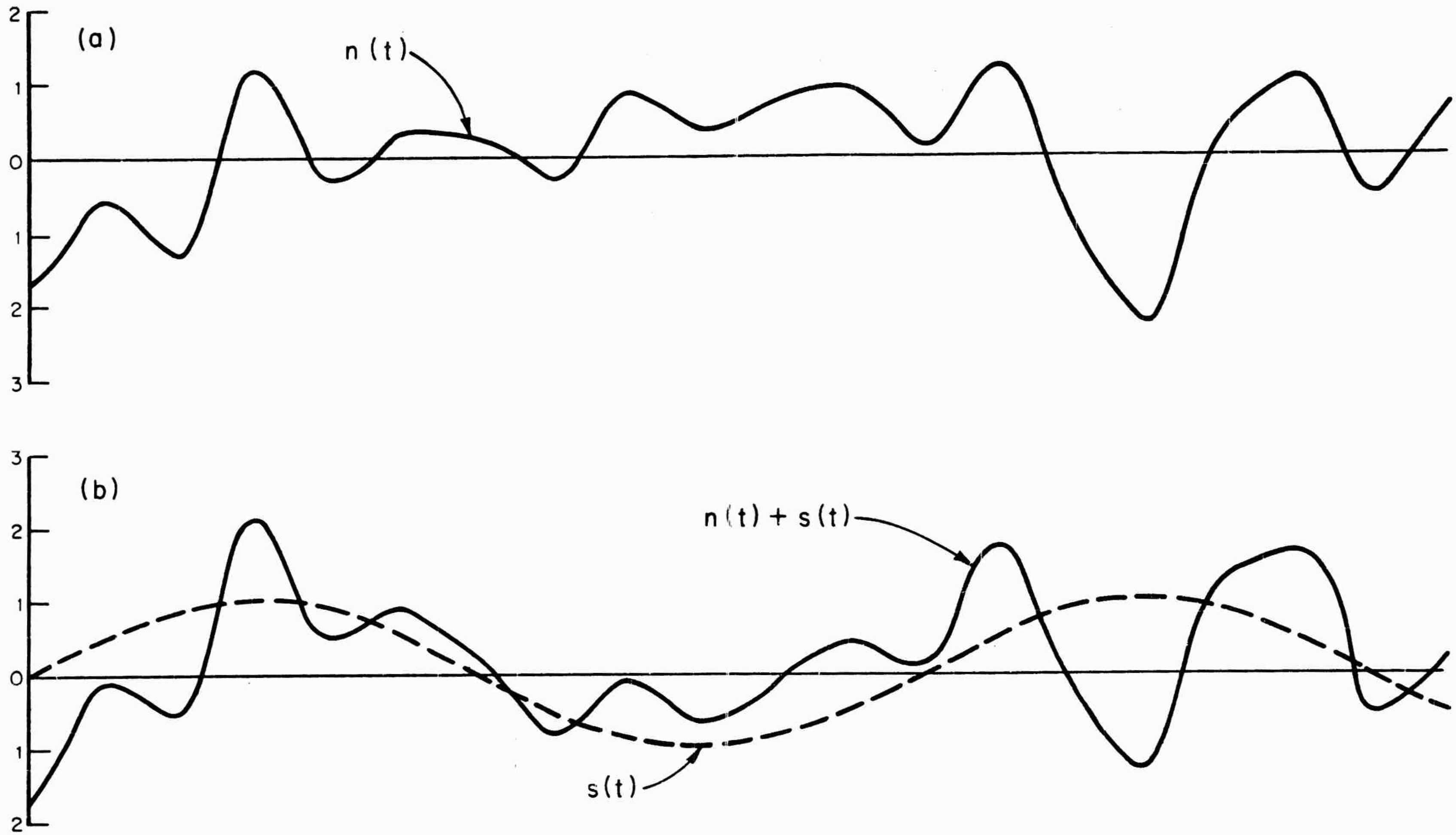


FIGURE 20 NOISE n(t) WITH AND WITHOUT A LOW-LEVEL SIGNAL s(t)

Now we introduce the geometrical concept of rotation of coordinates. The probability distribution of our observations is spherically symmetrical with respect to their centers, and hence retains the same form with a rotation of axes, i.e., the probability distribution of the new coordinate will still be normal with variance N/2 regardless of the new directions of the axes.

For skeptics, we shall illustrate these concepts for the simplest nontrivial case, two dimensions. Suppose x and y are given, statistically independent, with normal distribution about 0 with variance N/2. Rotate the coordinate axes by an angle θ. Then

$$u = x \cos\theta + y \sin\theta$$

$$v = -x \sin\theta + y \cos\theta$$

Let us look now at the mean* and variance of u:

$$\overline{u} = \overline{x \cos\theta + y \sin\theta} = \overline{x} \cos\theta + \overline{y} \sin\theta = 0$$

$$\overline{u^2} = \overline{(x \cos\theta + y \sin\theta)^2} = \overline{x^2 \cos^2\theta + 2\,x\,y \sin\theta \cos\theta + y^2 \sin^2\theta}$$

$$= (N/2) \cos^2\theta + (N/2) \sin^2\theta + \overline{xy} \cdot 2 \sin\theta \cos\theta$$

Note that the assumption that x and y are independent means simply that $\overline{xy} = \overline{x}\,\overline{y}$, which implies $\overline{xy} = 0$. Hence

$$\overline{u^2} = N/2$$

$$s^2 = \overline{u^2} - \overline{u}^2 = N/2$$

Similarly, the variance of v is N/2. Finally, u and v are statistically independent, for

$$\overline{u\,v} = \overline{(x \cos\theta + y \sin\theta)\ (-x \sin\theta + y \cos\theta)}$$

$$= (\overline{-x^2} + \overline{y^2}) \sin\theta \cos\theta = 0$$

---

*A horizontal bar over an expression signifies an average taken over a suitable range, usually an average over the statistical ensemble or a time average. Under a wide range of circumstances of interest (those satisfying ergodic conditions), the ensemble average and the time average are equal.

Now let us choose a new set of coordinates so that one of the axes is parallel to s(t). The representation of s(t) in the new coordinate system will consist of one term

$$s(t) = \sqrt{S} \quad \psi_1$$

so that obviously

$$\psi_1 = \frac{1}{\sqrt{S}} \quad s(t).$$

The noise is represented by

$$n(t) = \sum_{x=1}^{2WT} n_k' \quad \psi_k$$

where the prime is used simply to distinguish coordinates in the new coordinate system.

Our problem is now that of distinguishing between

$$f(t) = s(t) + n(t) = (\sqrt{s} + n_1') \psi_1 + \sum_{2}^{2TW} n_k' \psi_k, \text{ signal present.}$$

$$= n(t) = n_1' \psi_1 + \sum_{2}^{2TW} n_k' \quad \psi_k, \text{ signal absent.}$$

Obviously, there is no point in examining any term but the first. We can isolate the coefficient of the first function, ψ_1, by using scalar products.

$$f_1 = f(t) \cdot \psi_1(t) = \int_{-\infty}^{\infty} f(t) \quad \psi_1(t) \; dt = \frac{1}{\sqrt{S}} \int_{-\infty}^{\infty} f(t) \quad s(t) \; dt$$

and test the hypothesis

$$f_1 = \sqrt{3} \quad + \quad n_1' \qquad \text{signal present}$$

against

$$f_1 = n_1' \qquad \text{signal absent}$$

We know that n_1' is normally distributed about zero with variance N/2 just like any among the original components n_k, for we assumed a pure rotation of the coordinate system (even though we never explicitly found the new coordinate system). The two distributions are illustrated in Figure 21. The problem is reduced to that of identifying the quantity $\sqrt{S}$ when perturbed by a noise with variance N/2. The ratio of the signal to the standard deviation of the noise is

$$d = \frac{\sqrt{S}}{\sqrt{N/2}} = \sqrt{\frac{2S}{N}}$$

For reliable detection d must be somewhat greater than unity. If the probability that s(t) will be present is about 50%, and the penalty for missing it when it is present (which we call miss) is the same as the penalty for detecting it when it is not present (which we call false alarm or FA), then we would probably put the threshold of detection near $1/2\sqrt{S}$. This makes the error probability the same for the two circumstances. They are shown in the first two columns of Table I. In a true search situation, we are searching for a "needle in a haystack," and the signal is expected to be absent nearly always. Cutting down the false alarm rate becomes an operational problem, and it is advantageous to raise the threshold. The table shows two examples.

In any case, a value of d of about 8 is needed, and we can say roughly

$$\sqrt{\frac{2S}{N}} \sim 8$$

$$\frac{2S}{N} \sim 64$$

$$\frac{S}{N} \sim 32$$

$$S \sim 32N$$

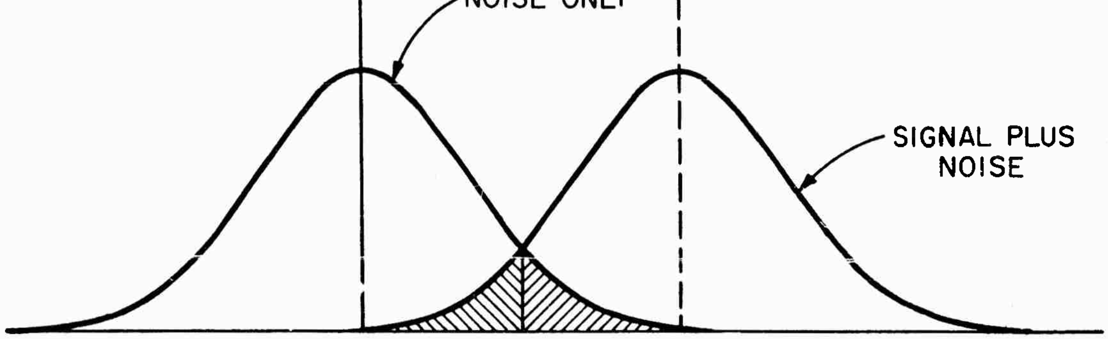


FIGURE 21 PROBABILITY DISTRIBUTION OF OUTPUT OF A COHERENT DETECTOR WHOSE INPUT IS WAVEFORMS LIKE THOSE IN FIGURE 20

TABLE I

PROBABILITY OF FALSE ALARM ERROR AND OF MISS ERROR AS A FUNCTION OF THRESHOLD LEVEL AND SIGNAL-TO-NOISE RATIO

Threshold	$1/2\sqrt{S}$		$\sqrt{S} - \sqrt{N/2}$		$\sqrt{S} - 2\sqrt{N/2}$	
$d = \sqrt{2S/N}$	FA	Miss	FA	Miss	FA	Miss
4	$2.3\cdot10^{-2}$	$2.3\cdot10^{-2}$	$1.4\cdot10^{-3}$	$1.6\cdot10^{-1}$	$2.3\cdot10^{-2}$	$2.3\cdot10^{-2}$
5	$6.2\cdot10^{-3}$	$6.2\cdot10^{-3}$	$3.2\cdot10^{-5}$	$1.6\cdot10^{-1}$	$1.4\cdot10^{-3}$	$2.3\cdot10^{-2}$
6	$1.4\cdot10^{-3}$	$1.4\cdot10^{-3}$	$2.9\cdot10^{-7}$	$1.6\cdot10^{-1}$	$3.2\cdot10^{-5}$	$2.3\cdot10^{-2}$
7	$2.3\cdot10^{-4}$	$2.3\cdot10^{-4}$	$2.0\cdot10^{-9}$	$1.6\cdot10^{-1}$	$2.9\cdot10^{-7}$	$2.3\cdot10^{-2}$
8	$3.2\cdot10^{-5}$	$3.2\cdot10^{-5}$	$2.6\cdot10^{-12}$	$1.6\cdot10^{-1}$	$2.0\cdot10^{-9}$	$2.3\cdot10^{-2}$

Let us compare that practical signal-to-noise ratio with the ideal case. Suppose we are concerned with a detection scheme in which there are 1,000,000 cells to look in. If we look and find something, then we have potentially distinguished among about 10^6 possibilities, and receive potentially about 20 bits. We shall, therefore, expect to need

$$S = 20 \times 0.693N = 13.9N$$

in signal energy.

However, there is error rate to consider. In a detection process, a rather liberal error rate is allowable, say $P(e) \sim .01$. Referring to the previously quoted formula

$$P(e) \sim 2^{-\nu\alpha C/R}$$

we recall that 2^{ν} is the number of binary digits constituting a message; by analogy, $\nu = 20$. Solving for R/C, one finds

$$R/C \sim .41$$

Hence, the amount of energy required in the signal to achieve an error rate of 0.01 is really

$$S = 20 \times 0.693\ N/.41 = 33N$$

This agrees very well with the value 32N derived above. The agreement is not fortuitous: this case fits the hypothesis of Fano's model quite precisely.

Notice that an error probability of 0.01 still requires a low false alarm rate: for the probability of a single false detection to be 0.01 in 10^6 cells, the probability of a false alarm in each cell must be less than 10^{-8}.

We see, therefore, that coherent detection, where viewed as a communication process, achieves about as much as one could expect. We need not look for new principles which will enable us to detect signals having less energy, but can devote ourselves to applying the conceptions of coherent detection and to engineering improvements to make the performance of such detectors live up to their design conception.

We can, of course, deliberately use a scheme like the one described above as a communication scheme. In such a case, it is usually impractical to search for one among a large number of signals. Costas* has described a system in which one of two signals, +s(t) or -s(t), is sent. Each one is "noise-like" in the sense of having no systematic pattern like a modulated carrier. This particular instance differs significantly from the model on which Fano's result is based: instead of sending energy S in one of 2^{ν} signals, one sends energy S/4 in all of them, with a resulting change in energy by a factor $2^{\nu-2}$. This is only an advantage if $\nu = 1$. However, for this particular case, which has a signalling rate R substantially less than the channel capacity C, the probability of error is

$$P(e) = \frac{2^{-C/R}}{2\sqrt{\pi \log 2\ C/R}}$$

The improvement in the exponent appears to be due to the use of +s(t) and -s(t), instead of +s(t) and 0, as the antithetical pair of signals; it does not seem likely that device will work where a choice among more than two signals is contemplated. For $P(e) = 10^{-6}$, $C/R \simeq 16$.

*See Reference 5 in the Bibliography.

IX. COHERENT AND INCOHERENT INTEGRATION

In a communication system, the alphabet of transmission signals will ordinarily be chosen so that one or another reasonably efficient demodulation process can be used. As we have seen, the efficiency of a detection system may be analyzed with the same mathematical tools. However, the designer of a detection system may not be able to control the signal or the environment enough to approach an optimum or efficient modulation scheme. An important example is that of a search process where the signal to be detected lasts a very long time, and where knowledge of its presence or absence is desired in a short time. This circumstance leadsto the idea of detection in a fixed or limited time, or, in the discrete case, of detection with a limited number of observations.

Heuristically, it is clear that increasing the observation time or the number of observations cannot decrease the certainty of detection, and should increase it. We are thus led to ask, how much is the detection process improved by increasing the observation time? We shall answer this question by suggesting a simple, plausible, and easily implemented criterion of effectiveness involving both observation time and signal-to-noise ratio, and show how increased observation time can be traded for decreased signal-to-noise ratio.

Suppose we have noise n(t) of bandwidth W, with a flat spectrum and rms amplitude N. Suppose we have a signal of constant d-c amplitude S. If the noise is present alone, the received waveform is

$$f(t) = n(t)$$

If the signal is present also, we have

$$f(t) = n(t) + S$$

An example of such signals is shown in Figure 22. In this example, N = 1, S = 1. We would like to test the following detection schemes

I. Correlation Detector: $\int_0^T \left[f(t)\right] S \, dt$

II. Square-law Detector: $\int_0^T f^2(t) \, dt$

III. Linear rectifier: $\int_0^T \left| f(t) \right| dt$

to find the relation among the signal-to-noise ratio S/N and the integration time T.

First, use the sampling theorem to characterize f(t) as a sequence

$$f_k = f(k/2W) \qquad k = 1, 2, ---, 2TW$$
$$n_k = n(k/2W)$$

Figure 22 shows how the samples are related to the continuous function f(t). The various samples f_k are independent and have a Gaussian distribution with variance N^2 (to avoid another proof, we can define this as Gaussian white noise of band-width W). To a high degree of approximation, we can replace the integrals (with appropriate constant multiplying factors) by sums:

$$\text{I.} \quad S_I = \sum_{1}^{2TW} f_k \simeq \frac{2W}{S} \int_0^T \left[f(t)\right] S\, dt$$

$$\text{II.} \quad S_{II} = \sum_{1}^{2TW} f_k^2 \simeq 2W \int_0^T f^2(t)\, dt$$

$$\text{III.} \quad S_{III} = \sum_{1}^{2TW} \left| f_k \right| \simeq 2W \int_0^T \left| f(t) \right| dt$$

We shall devote the rest of the discussion to the sums S_I, S_{II}, and S_{III}, and try to see how they depend on the integration-time and signal-to-noise ratio. The constant 2W and 2W/S is a scale factor and is not important for the present discussion.

Figure 23 shows the samples f_k, the squares of the samples f_k^2, and the absolute values of the samples $\left| f_k \right|$ for the noise, with and without signal, of Figure 22.

If we look at the signal f(t) at any instant; i.e., if we look at a single sample f_k, it has

mean value	$\mu_s = S$	if signal present
	$\mu_o = 0$	if signal absent
variance	$\sigma_s^2 = N^2$	if signal present
	$\sigma_o^2 = N^2$	if signal absent

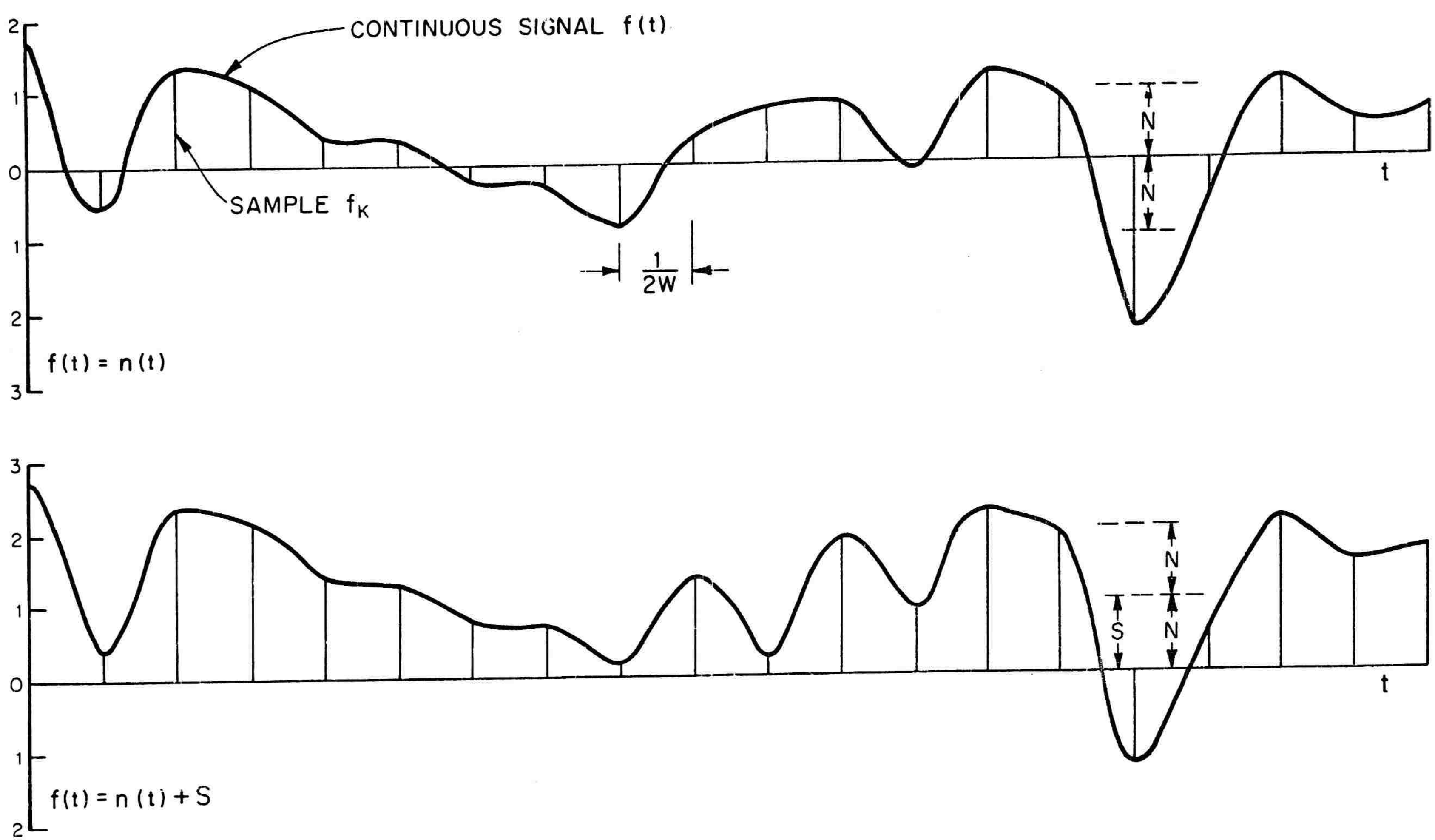


FIGURE 22 RANDOM NOISE n(t) WITH AND WITHOUT SUPERIMPOSED SIGNAL s(t) = S, SHOWING SAMPLES

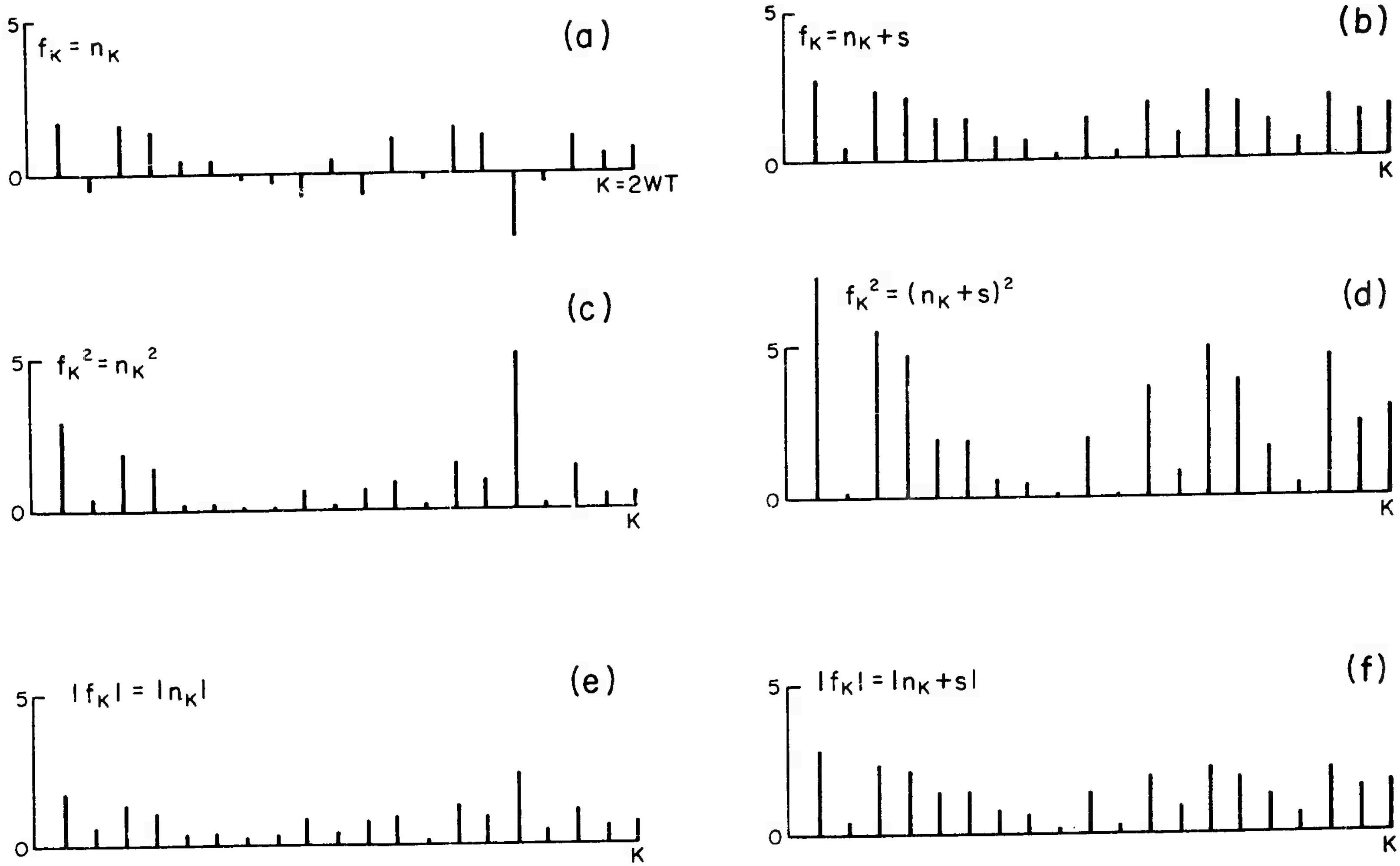


FIGURE 23 SAMPLES, SQUARES OF SAMPLES, AND ABSOLUTE VALUES OF SAMPLES IN THE ABSENCE AND IN THE PRESENCE OF SIGNALS

The two variances are the same, and we can ignore the distinction implied by the subscript. We can use $(\mu_S - \mu_O)/\sigma$ as a measure of effectiveness of a detection process. For a single sample of the signal, this is just the signal-to-noise ratio S/N. Figure 23a and b shows the samples f_k for the noise and for the signal plus noise shown in Figure 22. Figure 24a shows the sum

$$S_I = \sum f_k = \sum (S + n_k)$$

as a function of 2TW. The expected value of S_I is μ_S = 2TWS, and its variance is

$$\sigma_S^2 = \overline{\left[\sum (S + n_k)\right]\left[\sum (S + n_j)\right]} - (2TWS)^2$$

$$= \sum \sum \overline{(S^2} + \overline{Sn_j} + \overline{Sn_k} + \overline{n_k n_j)} - (2TWS)^2$$

$$= (2TW)^2 S^2 + 0 + 0 + \sum \overline{n_k^2} - (2TWS)^2$$

$$= \sum \overline{n_k^2}$$

$$= 2TWN^2$$

Repeating the computation with S = 0, we find that the mean value and variance of $\sum n_k$ are $\mu_o = 0$ and $\sigma_o = 2TWN^2$.

Figure 25a and b illustrates the distribution of observations to be expected after integration over a time such that 2TW = 20 and 100 respectively. The observations are distributed according to the well-known bell-shaped normal probability distribution. The center of the normal distribution curve is at the expected value μ, and the standard deviation is σ, the square root of the variance. In order to make an effective detector, it is necessary to set a threshold somehow

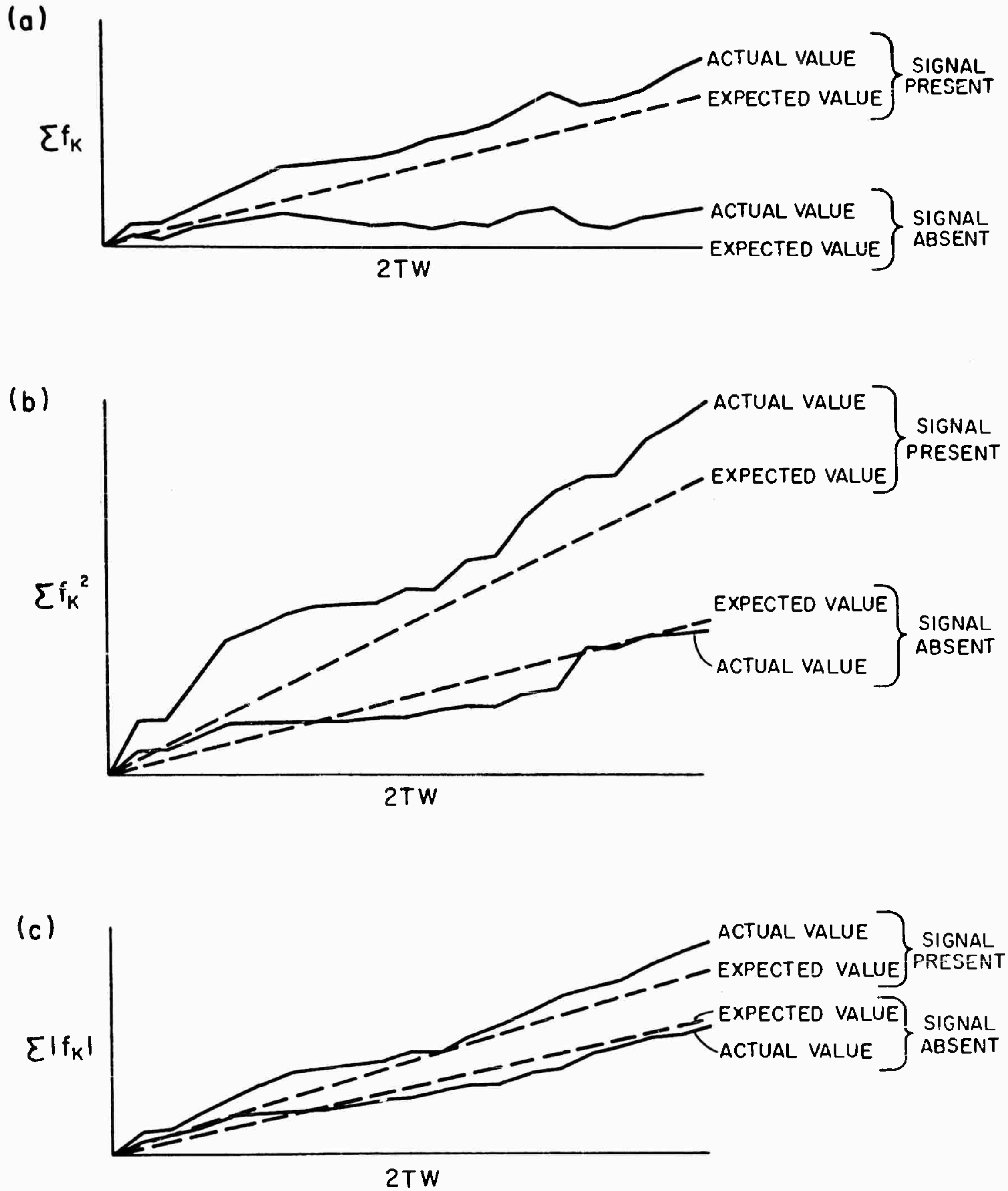


FIGURE 24 $\sum f_k$, $\sum f_k^2$, AND $\sum |f_k|$ IN THE ABSENCE AND IN THE PRESENCE OF SIGNALS, COMPARED WITH EXPECTED VALUES

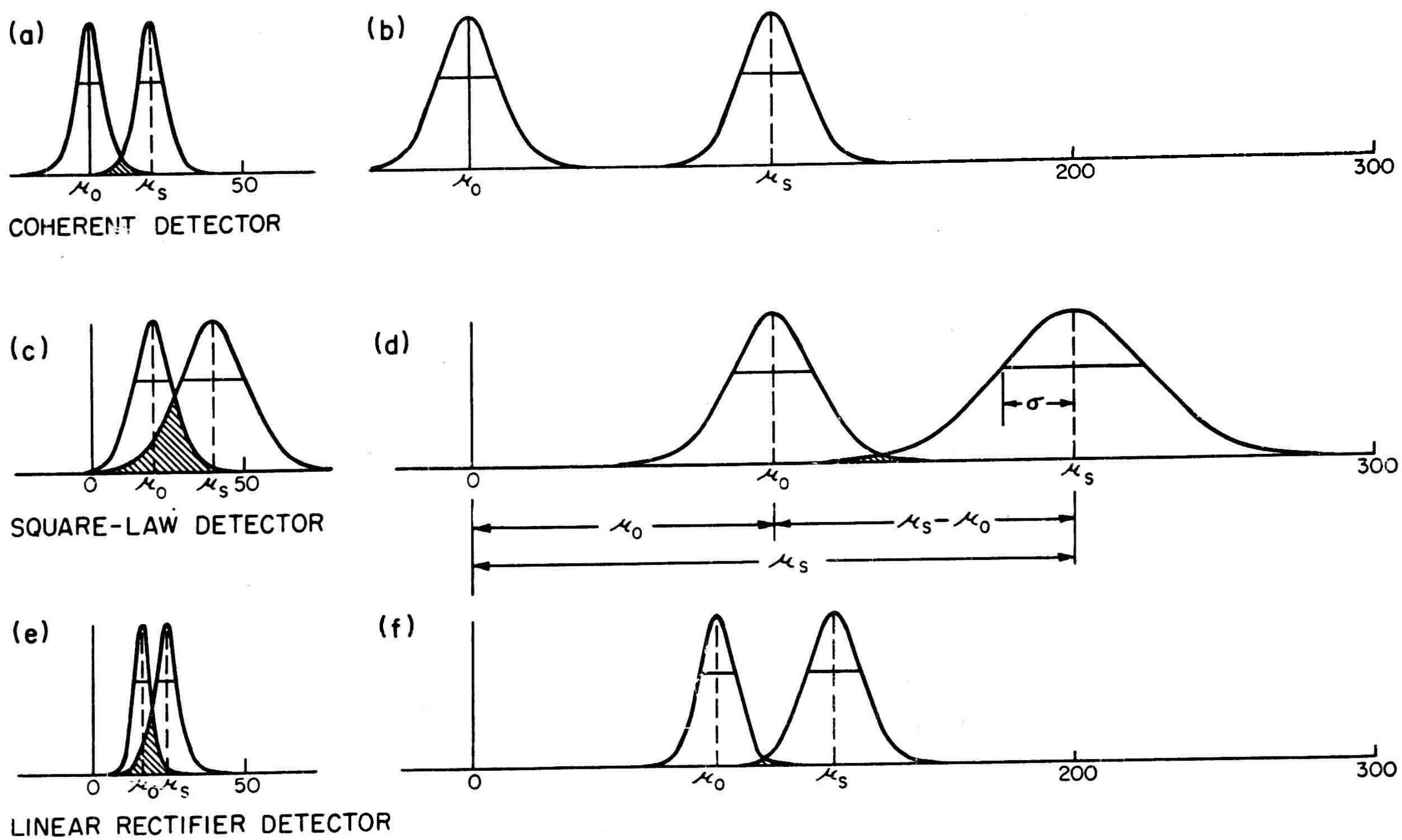


FIGURE 25 DISTRIBUTION OF OBSERVATIONS FOR COHERENT, SQUARE-LAW, AND LINEAR RECTIFIER DETECTION FOR TWO DIFFERENT INTEGRATION TIMES
(2TW = 20 for (a), (c), and (e); 2TW = 100 for (b), (d), and (f)

so that an observation will nearly always fall on one side of the threshold when the signal is present, and nearly always fall on the other side of the threshold where the signal is absent. Inasmuch as the probability distributions overlap (the overlapping part is cross-hatched in the figure), there is no place to establish a threshold which will give error-free results. A reasonable and useful measure of the effectiveness is the ratio of the distance between the peaks, $\mu_S - \mu_O$, and the width of the peak, measured by σ. When only one sample is taken, we have seen that the ratio is S/N. When 2TW samples are integrated in a coherent detector, the measure of effectiveness is

$$\frac{\mu_S - \mu_O}{\sigma} = \frac{2TWS - 0}{\sqrt{2TWN^2}} = \sqrt{2TW}\,\frac{S}{N}$$

i.e., the effect of integration over time T is equal to the effect of improving the S/N ratio by a factor $\sqrt{2TW}$.

In a square-law detector, the sample is squared to get $f_k^2 = n_k^2$ or $(n_k + S)^2$ (Figure 23c and d). Let us examine the build-up of the sum,

$$S_{II} = \sum f_k^2 = \sum (S + n_k)^2 = \sum (S^2 + 2Sn_k + n_k^2) \text{ (signal present)}$$

$$\text{or } = \sum n_k^2 \qquad \text{(signal absent)}$$

Its expected value is $2TWS^2 + 2TWN^2$ if the signal is present, and

$2TWN^2$ if the signal is absent

Figure 24b shows the actual growth of $\sum f_k^2$ compared with the expected value, for both cases.

We could find the variance of S_{II} by brute force. However, it is easier to work indirectly, and to define a new population whose members are

$$z_k = S^2 + 2Sn_k + n_k^2$$

and find sample mean and sample variance, and work indirectly to the sums.

If we define

$$n_k = N\,x_k$$

then x_k forms a normal population of variance unity, with a probability distribution

$$P(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \qquad \int_{-\infty}^{\infty} P(x)dx = 1$$

The mean value of x_k^2 is

$$\overline{x_k^2} = \int_{-\infty}^{\infty} x^2 P(x)\, dx = 1$$

The mean value of x_k^4 is

$$\overline{x_k^4} = \int_{-\infty}^{\infty} x^4 P(x)\, dx = 3$$

The mean values of x and x^3 are zero, because P(x) is a symmetric (even) function. For n_k, the means are

$$\overline{n_k^2} = N^2$$

$$\overline{n_k^4} = 3N^4$$

Now for

$$S^2 + 2S\, n_k + n_k^2$$

the mean value is

$$S^2 + 2S\, \overline{n_k} + \overline{n_k^2}$$

$$= S^2 + 0 + N^2$$

and the expected value of the sum is

$$\mu_S = \overline{\sum (s + n_k)^2} = 2TW\,(S^2 + N^2)$$

The variance of $(S + n_k)^2$ is

$$\overline{\left[\sum (S+n_k)^2\right]^2} - \overline{(S+n_k)^2}^2$$

$$= \overline{S^4 + 4S^3 n_k + 6S^2 n_k^2 + 4S n_k^3 + n_k^4} - (S^2 + N^2)^2$$

$$= S^4 + 0 + 6S^2N^2 + 0 + 3N^4 - S^4 - 2S^2N^2 - N^4$$

$$= 4S^2N^2 + 2N^4$$

Hence the variance of the sum is

$$\sigma_S^2 = \overline{\sum (S+n_k^2)} = 2TW\,(4S^2N^2 + 2N^4)$$

By repeating the computation with $S = 0$, we can find

$$\mu_o = 2TWN^2$$

$$\sigma_o^2 = 2TWN^4$$

Figure 25c and d shows normal distribution curves with these means and variances for $2TW = 20$ and 100.

An aggravating factor here is that the variance is different when the signal is present than it is when the signal is absent. Let us agree that we are most interested in the case $S/N << 1$. Then

$$2TW\,(4S^2N^2 + 2N^4) \simeq 2TW \cdot 2N^4 = 4TWN^4$$

independent of whether the signal is present.

Using the same criterion as before, we measure the effectiveness of the detector by

$$\frac{\mu_s - \mu_o}{\sigma} = \frac{2TW\,\left[(S^2 + N^2) - (N^2)\right]}{\sqrt{4TWN^4}} = \left(\frac{S}{N}\sqrt[4]{TW}\right)^2$$

i.e., integrating a time T is equivalent to improving S/N by a factor $\sqrt[4]{TW}$.

Now let us look at a rectifier. The samples are rectified to get $|f_k| = |n_k|$ or $|n_k + S|$ (Figure 23c and f), and the detector output after integration is

$$S_{III} = \sum |f_k|$$

Once again we examine the individual terms of the sum, and ask, what are the mean and variance of $|f_k|$?

$$\overline{|f_k|} = \overline{|S + n_k|} = \overline{|S + Nx_k|}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |S + Nx| \, e^{-x^2/2} \, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x=S/N} - (S + Nx) \, e^{-x^2/2} \, dx + \frac{1}{\sqrt{2\pi}} \int_{-S/N}^{\infty} (S + Nx) e^{-x^2/2} dx$$

Here we can evaluate the integral approximately by a tedious but straightforward process, as follows:

$$\text{Substitute} \int_{-\infty}^{0} + \int_{0}^{S/N} \text{ for } \int_{0}^{S/N} \text{ and analogously} \int_{S/N}^{0} + \int_{0}^{\infty} \text{ for } \int_{S/N}^{\infty}$$

Evaluate all integrals in $(0, \infty)$ and $(\infty, 0)$ exactly. Evaluate integrals in $(-S/N, 0)$ and $(0, S/N)$ by using the approximation

$$e^{-x^2/2} \simeq 1$$

The result is

$$\overline{|f_k|} \simeq \frac{2N}{\sqrt{2\pi}} \left(1 + \frac{S^2}{2N^2}\right) \quad (S/N) < 1$$

The expected value of the sum is

$$\mu_S = \overline{\sum |S + n_k|} \simeq \frac{2TW}{\sqrt{2\pi}} \left(2N + \frac{S^2}{N}\right)$$

when the signal is present, and

$$\mu_o = \overline{\sum |n_k|} \simeq \frac{2TW}{2\pi} \quad 2N$$

when the signal is absent. The difference is

$$\mu_s - \mu_o = \frac{2TW}{\sqrt{2\pi}} \cdot \frac{S^2}{N}$$

What about the variance? The mean square value is easy to evaluate; for the absolute value operation is trivial when the function is squared:

$$\overline{|S + n_k|^2} = \overline{(S + n_k)^2} = S^2 + N^2$$

One must be careful not to jump to conclusions, however. The mean value laboriously computed above must now be used.

$$\text{var}\left\{|S + n_k|^2\right\} = \overline{|S + n_k|^2} - \left[\overline{|S + n_k|}\right]^2$$

$$\simeq S^2 + N^2 - \frac{4N^2}{2\pi} - \frac{4S^2}{2\pi} - \frac{1}{2\pi} \frac{S^4}{N^2}$$

$$\sigma_s^2 \sim 2TW \left[S^2 + N^2 - \frac{2N^2}{\pi} - \frac{2S^2}{\pi} - \frac{S^4}{2\pi N^2}\right]$$

If S/N < < 1, this is approximately

$$\sigma_s^2 \simeq 2TWN^2 \left(1 - \frac{2}{\pi}\right)$$

Similarly

$$\sigma_o^2 \simeq 2TWN^2 \left(1 - \frac{2}{\pi}\right)$$

These probability distributions are plotted in Figure 25e and f for 2TW = 20 and 100.

The measure of merit of the detector is

$$\frac{\mu_s - \mu_o}{\sigma} = \frac{2TWS \cdot \frac{S}{N}}{\sqrt{2\pi} \cdot \sqrt{2TW\left(1 - \frac{2}{\pi}\right)} \cdot N}$$

$$= \left(\frac{S}{N} \quad \sqrt[4]{\frac{TW}{\pi - 2}}\right)^2$$

i.e., the effect of integration for a time T is equivalent to the effect of improving

$$\frac{S}{N} \text{ by a factor } \sqrt[4]{\frac{TW}{\pi - 2}}$$

Note that this is just a shade worse than $\sqrt[4]{TW}$.

The ratio is $\sqrt[4]{\frac{1.00}{1.1416}}$ or approximately 0.1 db.

Table II summarizes the expected values and variances of the outputs of these three kinds of detectors. The effect of integration with a coherent detector over a time T is equivalent to an improvement in the input signal-to-noise ratio of a factor $\sqrt{2TW}$. This is sometimes stated as 3 db improvement per doubling of integration time. The effect of integration with an incoherent square-law or linear rectifier detector over a time T is equivalent (when the input S/N is low) to an improvement in the signal-to-noise ratio of $\sqrt[4]{TW}$ or $\sqrt[4]{TW/(\pi - 2)}$ respectively. This is sometimes stated as 1.5 db improvement per doubling of integration time.

There is another respect in which the square-law and linear rectifier detectors are inferior to the coherent detector. The distributions in Figure 25 and in Table II show that the expected value of the output of a coherent detector depends on the signal only, and the variances on the noise only, whereas in square-law and linear rectifier detectors the expected values and variances depend jointly on signal and noise. Now to a first approximation, the best place to put the detection threshold depends on the expected value of the output, and not on the variance. This means that the threshold can be set in a coherent detector independent of the noise. This is not possible in linear or square-law detectors, for the output wanders back and forth as the noise level varies. Unless the noise is very uniform, as, for example, is thermal noise in a low-noise electronic amplifier, some extra provision must be made to compensate for secular variations in noise level.

TABLE II

EXPECTED VALUE AND VARIANCE OF THE OUTPUTS OF SEVERAL TYPES OF DETECTORS

	Noise Only		Signal Plus Noise	
	Expected Value	Variance	Expected Value	Variance
Coherent Detector	0	$2TWN^2$	$2TWS$	$2TWN^2$
Square-Law Detector	$2TWN^2$	$4TWN^2$	$2TW(S^2 + N^2)$	$2TW(4S^2 N^2 + 2N^4)$
Linear Rectifier Detector	$\frac{4TWN}{\sqrt{2\pi}}$	$2TWN^2(1 - \frac{2}{\pi})$	$\frac{2TW}{\sqrt{2\pi}}(2N + \frac{S^2}{N})$ or	$2TW\left[(S^2 + N^2)(1 - \frac{2}{\pi}) - \frac{S^4}{2\pi N^2}\right]$

These results were derived for a very particular signal waveform, a rectangular d-c pulse. The conclusions are quite generally valid, however. The restriction to low input S/N is relatively unimportant in most practical cases, for the output signal-to-noise ratio in all three of these detectors, and we can concentrate our attention on the "worst case," where the signal-to-noise ratio is as low as the system can stand.

What is the difference between coherent and incoherent detection? In the geometric language in which we represent each of a family of signals by a point in a space of 2WT dimensions, coherent detection makes use of the direction of the point relative to the coordinate axes as well as the distance, whereas incoherent detection uses the distance only.

Is there any detection which is "intermediate" between coherent and incoherent? Such systems have been described by Jacobs* and others. In the system described by Jacobs, a band of the spectrum is divided into a number of discrete equal. bands. The signal is made up of bursts of energy, not overlapping in time, and each is confined to one of the bands. Within each band, the energy is detected incoherently.

Let us examine why this is partly "coherent." Suppose the time duration of a burst is T, the total bandwidth is W, and the bandwidth of each of k bands is W/k = B. Let us imagine the signal represented not by its amplitude samples but by its frequency components.

$$f(t) = \sum_{n=1}^{TW} \left[a_n \cos (2 \pi n t/T) + b_n \sin (2\pi n t/T) \right]$$

(For convenience, it is assumed that the signal lies in the band of frequency from 0 to W, but it could lie elsewhere with appropriate changes in representation.) The coefficients a_n and b_n are the coordinates, and the number of coordinates is 2TW (give or take a few, depending on whether we assume a d-c term and whether TW is an integer or not).

Now let us look at a signal falling in a particular band, say

$$m B < f < (m + 1) B$$

---

* "Optimum Integration Time for the Incoherent Detection of Noise-like Communication Signals," I. Jacobs, Bell Telephone Laboratories, Inc., Whippany, N.J., Presented at the 1962 Spring URSI Meeting, April 30 - May 3.

This is representable by

$$f(t) = \sum_{n=mBT+1}^{(m+1)BT} \left[(a_n \cos (2\pi nt/T) + b_n \sin (2\pi nt/T) \right]$$

involving only 2BT terms. The receiver filters the incoming signal into a band $mB < f \leq (m+1)B$, and hence makes use of the fact that all components of f(t) lie in a given subset of the possible directions. But after filtering, it uses an incoherent detector which makes no further use of the detailed relations among the components.

When the parameters are duly proportioned, this modulation and detection scheme is reasonaly efficient. In the band in which it falls, the transmitted signal should have a spectral power density about three times that of the noise for most efficient transmission. For most efficient performance, the number of bands, k, should be hundreds, and the information transmitted per burst is $\log_2 k$. The burst length is of the order of magnitude 20/B, and the optimum is more or less dependent on the number of bands, k. The amount of power required per bit is around 60 (.693 N) for k = 2 and falls to around 10(.693 N) for k of several hundreds. On this basis, it is competitive with AM, SSB, and FM, and not much worse than FM with feedback.

Why would such a modulation scheme be used? The detailed signal structure required for coherent detection is destroyed or degraded by such phenomena as doppler shift, which obscures small frequency shifts, or multipath propagation, which destroys small time distinctions. With a signal in a band of total bandwidth kB and time duration 20/B, we should require frequency discrimination approximating B/20 or time discrimination approximating 1/kB to make coherent detection possible, whereas this system operates with much coarser frequency bands of bandwidth B and much coarser time segments of length 20/B. In round numbers, its frequency discrimination is 10 times coarser or its time discrimination 1000 times coarser than those required by coherent detection schemes depending exclusively on frequency discrimination or time discrimination, respectively.

X. CONCLUSION

Where, now, has this comparison of modulation and detection systems brought us? It has been shown that there is a minimum average energy required to transmit one bit of information in the presence of random noise of fixed intensity and uniform spectral distribution. The degree to which amplitude modulation, single-sideband modulation, frequency modulation, frequency modulation with feedback, and a particular frequency-band-limited noise-pulse modulation system approach the ideal has been estimated, and all were found to require three to 100 times more energy per bit than the ideal minimum. Detection of a signal in a noisy background, as in a radar, was viewed as a communication process, and it was found that the energy required per bit of effective information received is only slightly more than the ideal minimum.

Implicitly, we have seen how to encode an information-carrying signal of relatively narrow bandwidth and high signal-to-noise ratio in a new form having broad bandwidth and low signal-to-noise ratio. When the formula for channel capacity was developed, it became obvious at once that channels having a high signal-to-noise ratio used more power than is necessary to transmit their information. On the other hand, for a communication channel to be useful to the ultimate users, the received message must have a relatively high message-to-noise ratio, that is, the error rate must be low. In all of the more straightforward and naive ways of modulating and demodulating, the signal is so much like the message that to keep a high message-to-noise ratio, we must have a high signal-to-noise ratio.

The derivation we gave of the channel-capacity formula suggests one relatively complex way to signal through a noisy channel without introducing errors into the message: by using almost countless numbers of noise-like waveforms as an alphabet of digital signals. This solution to the problem is conceptually easy to handle, and on paper allows us to reach significant results. However, everyone seems to agree that this is an undesirable way to modulate and demodulate, or to code and decode, because it would require extremely complex equipment. But now frequency-modulation-with-feedback is a way of making a trade among bandwidth, power, and signal-to-noise ratio which does not involve resorting to complicated digital codes.

Other advantages besides saving of transmitter power arise from the efficient use of a communication channel. For example, if we consider the efficient utilization of space in our signal-space of 2WT dimensions, we realize that in signal-space any noise is as good as any signal, and no signal is any better than any noise. Thus, we find that in such a context it is impossible to have especially obnoxious jamming signals. There is no more efficient signal

for jamming than random noise, and we already know that under these circumstances the system can be designed to operate with a very low signal-to-noise ratio. We can see that to jam such a system successfully, we must put into the receiver more jamming power than signal power. This makes jamming costly.

There is another benefit from operating with a very low signal-to-noise ratio. If we can really work a communications system so that the signal power-level is much lower than the noise power-level, we introduce the possibility of signalling in such a way that it is hard to tell whether any signal is being transmitted at all. We can thus indirectly make the jamming problem more difficult again, for the jammer must first hunt around to find out where there is something to jam before he knows whether to waste his effort trying to jam it.

By looking at seaching for the presence of a signal as a communication process, we have learned that there is a limit to the detectibility of a single signal in a noise background, and that this limit is described in terms of the noise energy density and the received signal energy. The shape of the signal wave is not significant as long as it is fully known in advance to the detector. The process of measuring the correlation between the known signal waveform and the received wave is known as coherent detection. If the signal waveshape is not completely known, certain kinds of incoherent detection, which vary according to the degree of ignorance of the signal waveshape, are possible. The less that is known about the signal waveshape, the more signal energy is required to assure positive detection. If the signal is a single pulse or a burst of a sinusoidal wave, one can use very simple detectors which approach the theoretical limit of search performance. Many signals which at first contact appear to be quite specific, such as, for example, the acoustic signal resulting from the spoken sound "ee," do in fact vary over a wide range, and are correspondingly hard to detect reliably.

In summary, the ultimate limit to the rate of transmission of information in a noisy background, or to the detection of a signal in a noisy background, is primarily determined by the noise power density and the signal power or energy. To approach this theoretical limit, the receiver must have precise detailed knowledge of the possible waveshapes of the transmitted signal. In the absence of such knowledge, more signalling power or energy is required.

BIBLIOGRAPHY

1. "Communication in the Presence of Noise," C. E. Shannon, Proc. I.R.E., Vol. 47, No. 1, January 1959, pp. 10-21.

2. "An Outline of Information Theory," E. N. Gilbert, The American Statistician, Vol. 12, February 1958, pp. 13-19.

3. "Interplanetary Communications," J. R. Pierce and C. C. Cutler, Advances in Space Science, Vol. 1, ed. by F. I. Ordway, III, Academic Press, 1959.

4. Transmission of Information, Robert M. Fano, MIT Press and John Wiley & Sons, New York, 1961.

5. "Poisson, Shannon and the Radio Amateur," J. P. Costas, Proc. I.R.E., Vol. 47, No. 12, pp. 2058-2068, December 1959.

6. "Prediction and Entropy of Printed English," C. E. Shannon, Bell System Technical Journal, Vol. 30 (1951), pp. 50-64.

DISTRIBUTION LIST

	No. of Copies
Bureau of Ships, Code 370 Department of the Navy Washington 25, D.C.	10
Chief of Naval Operations (OP-71) Department of the Navy Washington 25, D.C.	10
Commanding Officer & Director U.S. Navy Underwater Sound Laboratory Fort Trumbull, New London, Conn. Attn: Mr. H. E. Nash	4
Office of Naval Research, Code 400X Department of the Navy Washington 25, D C.	2
Director U. S. Naval Research Laboratory Washington 25, D.C. Attn: Dr. H. Saxton	2
Chief, Bureau of Naval Weapons Department of the Navy Washington 25, D.C. Attn: Mr. I. H. Gatzke	2
Commanding Officer U.S. Naval Air Development Center Johnsville, Pa. Attn: Mr. R. I. Mason	2
Commanding Officer & Director U S. Navy Electronics Laboratory San Diego 52, California Attn: Dr. D. A. Wilson	2

	No. of Copies
Commander, Antisubmarine Warfare Force U. S. Pacific Fleet Navy No. 128, c/o FPO San Francisco, California	2
Marine Physical Laboratory of the Scripps Institution of Oceanography San Diego 52, California	1
Technical Library (Code P80962) U.S. Naval Ordnance Test Station Pasadena Annex 3202 East Foothill Blvd. Pasadena 8, California	1
Officer in Charge U.S. Naval Oceanographic Unit Building #1, Treasure Island San Francisco 30, California	1
Commander Oceanographic System Atlantic Norfolk 11, Virginia	1
Commander, Antisubmarine Warfare Force U.S. Atlantic Fleet Norfolk 11, Virginia	1
Commanding Officer U. S. Navy Underwater Sound Laboratory Research Detachment Navy No. 138, c/o FPO New York, N. Y. Attn: Dr. W. A. VonWinkle	1
Hudson Laboratories P. O. Box 239 Dobbs Ferry, N. Y. Attn: Dr. R. A. Frosch	1

	No. of Copies
President Naval War College Newport, Rhode Island	1
Commander Destroyer Development Group Two U. S. Atlantic Fleet Newport, Rhode Island	1
Director of Research Institute of Naval Studies 185 Alewife Brook Parkway Cambridge 38, Mass.	1
Director, ASW Systems Analysis Group Undersea Warfare R&D Planning Council U. S. Naval Ordnance Laboratory White Oak, Silver Spring, Md.	1

PROJECT TRIDENT TECHNICAL REPORTS

Report No.	Title
1011260	COLOSSUS I, December, 1960 (C)
1021260	THEORETICAL INVESTIGATION OF CROSS-FIX PROBLEMS AND CORRELATION EFFECTS, December. 1960 (C)
1031260	THE SUBMARINE AS A SURVEILLANCE PLATFORM, December, 1960 (S)
1041260	Title Classified, December, 1960 (S)
1051260	AIRBORNE JEZEBEL, December, 1960 (S)
1061260	SURFACE-SHIP SONARS IN OCEAN-AREA SURVEILLANCE, December, 1960 (S)
1071260	Title Classified, December, 1960 (S)
1080361	LOW-POWER ENERGY SOURCES, March, 1961 (C)
1090561	SOLUS, May, 1961 (C)
1100561	NONACOUSTIC METHODS FOR SUBMARINE DETECTION, May, 1961 (S)
1110561	ARTEMIS, May, 1961 (S)
1120561	COLOSSUS II, May, 1961 (S)
1130961	RELIABILITY OF UNATTENDED ELECTRONICS EQUIPMENT, September, 1961 (U)
1141061	NUTMEG, October, 1961 (S)
1150162	DEEP JULIE, January, 1962 (S)
1160262	METHODS FOR ANALYZING THE PERFORMANCE OF DISTRIBUTED FIELDS OF DETECTORS, February, 1962 (C)
1170262	MAGNETIC ANOMALY DETECTORS IN FIXED SHALLOW WATER BARRIERS, February, 1962 (S)

Report No.	Title
1180262	ELECTRICAL CONDUCTIVITY, COMPRESSIBILITY, AND VISCOSITY OF AQUEOUS ELECTROLYTIC SOLUTIONS, February, 1962 (U)
1190462	A FEASIBILITY STUDY OF THE PASSIVE DETECTION OF QUIET SUBMARINES, April, 1962 (S)
1200562	Title Classified, May, 1962 (S)
1210562	RADIATED NOISE CHARACTERISTICS OF DIESEL-ELECTRIC SUBMARINES, May, 1962 (S)
1220562	DIRECTIVE RECEIVING ARRAYS, May, 1962 (C)
1230662	ANALYTICAL BACKGROUND OF COMPUTATIONAL METHODS FOR UNDERWATER SOUND PROPAGATION, June, 1962 (U)
1240762	MARINE CORROSION AND FOULING, July, 1962 (U)
1250862	SURVEY ON AMBIENT SEA NOISE, August, 1962 (C)
1260862	DEEP SUBMERSIBLE WORK VEHICLES, August, 1962 (C)
1270862	THE EFFECT OF PRESSURE ON THE ELECTRICAL CONDUCTIVITY OF SEA WATER, August, 1962 (U)
1281262	ENGINEERING PROPERTIES OF MARINE SEDIMENTS, December, 1962 (U)

AD ______________ Accession No. ______________
Arthur D. Little, Inc.
Cambridge, Massachusetts

UNCLASSIFIED

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

(Project Trident Tech. Rept. No. 1291262) Dec. 1962, pp. 80
Illus. 25

(Contract NObsr 81564)

This is an expository essay on information theory for engineers interested in communications, sonar, and radar who have no specialized knowledge of statistical communication theory. The fundamental concepts of information theory, and in particular, quantity of information and channel capacity, are defined and explained in simple terms. These concepts are used to make a quantitative estimate of the performance of several common modulation schemes and to analyze the performance of search and detection systems. The effectiveness of repeated or prolonged observations on detection thresholds and reliability of detection, and the relative performance of coherent and incoherent integration, are explained and illustrated quantitatively.

UNCLASSIFIED

AD ______________ Accession No. ______________
Arthur D. Little, Inc.
Cambridge, Massachusetts

UNCLASSIFIED

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

(Project Trident Tech. Rept. No. 1291262) Dec. 1962, pp. 80
Illus. 25

Contract NObsr 81564

This is an expository essay on information theory for engineers interested in communications, sonar, and radar who have no specialized knowledge of statistical communication theory. The fundamental concepts of information theory, and in particular, quantity of information and channel capacity, are defined and explained in simple terms. These concepts are used to make a quantitative estimate of the performance of several common modulation schemes and to analyze the performance of search and detection systems. The effectiveness of repeated or prolonged observations on detection thresholds and reliability of detection, and the relative performance of coherent and incoherent integration, are explained and illustrated quantitatively.

UNCLASSIFIED

AD ______________ Accession No. ______________
Arthur D. Little, Inc.
Cambridge, Massachusetts

UNCLASSIFIED

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

(Project Trident Tech. Rept. No. 1291262) Dec. 1962, pp. 80
Illus. 25

(Contract NObsr 81564)

This is an expository essay on information theory for engineers interested in communications, sonar, and radar who have no specialized knowledge of statistical communication theory. The fundamental concepts of information theory, and in particular, quantity of information and channel capacity, are defined and explained in simple terms. These concepts are used to make a quantitative estimate of the performance of several common modulation schemes and to analyze the performance of search and detection systems. The effectiveness of repeated or prolonged observations on detection thresholds and reliability of detection, and the relative performance of coherent and incoherent integration, are explained and illustrated quantitatively.

UNCLASSIFIED

AD ______________ Accession No. ______________
Arthur D. Little, Inc.
Cambridge, Massachusetts

UNCLASSIFIED

AN INTRODUCTION TO MODULATION, CODING, INFORMATION THEORY, AND DETECTION

(Project Trident Tech. Rept. No. 1291262) Dec. 1962, [illegible]
Illus. 25

(Contract NObsr 81564)

This is an expository essay on information theory for engineers interested in communications, sonar, and radar who have no specialized knowledge of statistical communication theory. The fundamental concepts of information theory, and in particular, quantity of information and channel capacity, are defined and explained in simple terms. These concepts are used to make a quantitative estimate of the performance of several common modulation schemes and to analyze the performance of search and detection systems. The effectiveness of repeated or prolonged observations on detection thresholds and reliability of detection, and the relative performance of coherent and incoherent integration, are explained and illustrated quantitatively.

UNCLASSIFIED